A Primer on Space-time Modeling from a Bayesian Perspective

LA-UR-05-3097

Dave Higdon

Statistical Sciences Group

Los Alamos National Laboratory

1 Introduction

The following notes are designed to get the interested reader up and running – developing his or her own spatial or space-time models – as quickly as possible. For better or worse, this means many details regarding alternative modeling, model checking, computation are omitted for the sake of keeping this primer streamlined. The goal of this primer is to convey an overall sense of what is involved in developing space-time models. Far more detailed accounts of spatial and space-time modeling can be found in standard texts such as Cressie (1991), Stein (1999), Wackernagel (1995), Chilés and Delfiner (1999), or Banerjee *et al.*(2004).

In this primer, the emphasis is on Gaussian spatial and space-time models. These models are quite flexible and can be adapted to wide variety of applications, even where the observed data are markedly non-Gaussian as will be shown in Section 4. Due to my own bias and experience, purely spatial models are first developed in Sections 2 through 4. These models are then extended to the space-time domain in Sections 5 and 6. Throughout the book, the computation for the actual applications is carried out using Markov chain Monte Carlo (MCMC), which was first implemented on a spatial

system in 1953. I'll do my best to show the natural links between spatial systems and MCMC. I also caution the reader, that for the sake of reducing the number of details that must be covered, MCMC is essentially the only method used in this primer. Though this technique is widely applicable, in many situations, there may be more efficient computational approaches that I neglect here.

2 Gaussian computation

The spatial models laid out in this primer make heavy use of Gaussian systems. This section goes over some basics for simulation and conditioning with multivariate normal models.

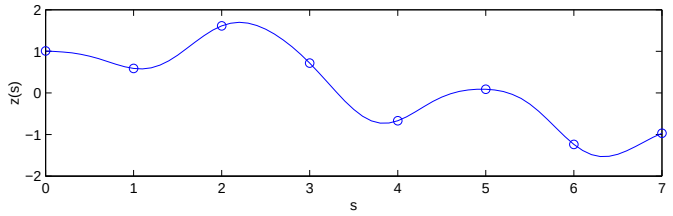


Figure 1: A realization from a 1-d Gaussian process (GP) with covariance given by $\Sigma_{ij} = \exp\{-||s_i - s_j||^2\}$.

Figure 1 shows a realization $z(s)$ of a Gaussian process model on $s_1, \ldots, s_n$

$$z = \begin{pmatrix} z(s_1) \\ \vdots \\ z(s_n) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \quad \Sigma \quad \end{pmatrix} \right)$$

where $\Sigma_{ij} = \exp\{-||s_i - s_j||^2\}$ and $||s_i - s_j||$ denotes the distance between locations s_i and s_j. Here z has a multivariate normal probability density function given by

$$\pi(z) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2} z^T \Sigma^{-1} z\}.$$

If the spatial locations $\{s_1, \ldots, s_n\}$ are taken to be $\{0, 1, \ldots, 7\}$ the realizations correspond to the circles plotted in Figure 1. Alternatively, the spatial locations $\{s_1, \ldots, s_n\}$ could define a very dense grid of points between 0 and 7. In this case, the realization corresponds to the line in Figure 1. Additional realizations of $z(s)$ are shown in Figure 2.

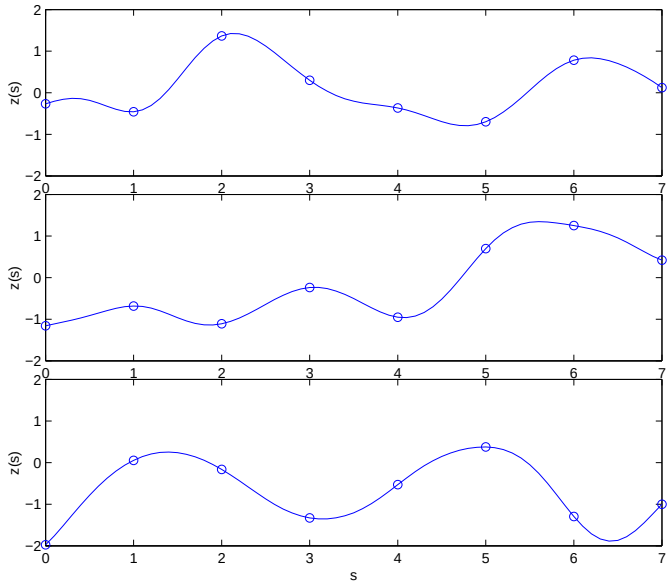


Figure 2: Several independent realizations from a 1-d Gaussian process with covariance given by $\Sigma_{ij} = \exp\{-||s_i - s_j||^2\}$.

The covariance matrix Σ for the process $z(s)$ is built according to the rule $\Sigma_{ij} = \exp\{-||s_i - s_j||^2\}$. In general, not just any rule will do – a covariance rule needs to give a valid covariance matrix (symmetric and positive definite) for any set of points within the spatial domain. For a catalog of possible covariance rules, consult previously mentioned texts on Kriging and spatial modeling. Later in this section, the covariance rules $\Sigma_{ij} = \exp\{-||s_i - s_j||^2\}$ and $\Sigma_{ij} =$

$\exp\{-||s_i - s_j||^1\}$ are compared. Generally, the smoothness and the strength of dependence in the realizations $z(s)$ are influenced by the choice of covariance rule. Note that practical covariance rules will need to acount for scaling in distance as well as variance. Thus a more general covariance rule might look like

$$\Sigma_{ij}(\sigma_z^2, r) = \sigma_z^2 \exp\left\{-(||s_i - s_j||/r)^2\right\}$$

where σ_z^2 controls the marginal varance of $z(s)$ and r scales distance. However, since the focus of this chapter through Section 3.2 is on $z(s)$, the scaling parameters are left at 1 for now.

2.1 Generating multivariate normal realizations

Most any computer package will generate independent, univariate normal draws:

$$u \sim N(0, I_n).$$

A standard property of normals –

$$\text{if } u \sim N(\mu, \Sigma), \quad \text{then } z = Ku \sim N(K\mu, K\Sigma K^T)$$

– can be used to construct correlated realizations z from iid draws held in vector u. The following recipe can be used to generate $z \sim N(0, \Sigma)$:

1. compute square root matrix L such that $LL^T = \Sigma$;
2. generate $u \sim N(0, I_n)$;
3. set $z = Lu \sim N(0, LI_nL^T = \Sigma)$.

This simple recipe gives some insight about how one might represent the process z in terms of basis vectors:

- the columns of L are effectively basis vectors for representing realizations:

$$z = \sum_{i=1}^{m} \ell_i u_i$$

  where ℓ_i is the ith column of L;

- the choice of L is not unique – alternative choices for L lead to different basis representations for z;

- L need not be square. If L is a $n \times m$ matrix, then the columns of L give a lower-dimensional basis representation for z if $m < n$, and an overspecified representation if $m > n$.

To get an idea of basis representations various square root factorizations produce, we consider a simple 1-d GP $z(s)$ at 20 spatial locations $\{s_1, \ldots, s_{20}\}$ which are equally spaced between 0 and 10 in ascending order. The covariance rule is given by $\Sigma_{ij} = \exp\{-||(s_i - s_j)/5||^2\}$. Hence the 20-vector $z = (z(s_1), \ldots, z(s_{20}))^T$ has a normal distribution with mean 0 and covariance Σ.

The most common method for constructing a square root L of a covariance matrix Σ is with the standard Cholesky factorization (Press *et al.*, 2002, Ch 2.9). The resulting columns of the square root matrix L are shown in Figure 3. Because the standard Cholesky factorization builds the basis vectors sequentially, without regard to the spatial locations $s_1, \ldots, s_{20}$ implicit in the vector z, the resulting basis representation is rather inefficient since each of the basis vectors have elements of appreciable size.

This is due to the ordering of the components of z that the resulting Cholesky decomposition of Σ is relatively inefficient. If one reorders the vector z so that its most correlated elements are not always adjacent, the resulting basis vectors

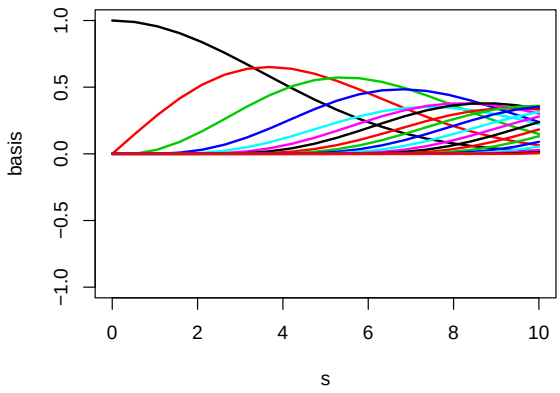


Figure 3: Basis vectors resulting from the Cholesky factorization of the covariance matrix obtained by applying the covariance rule $\Sigma_{ij} = \exp\{-||(s_i - s_j)/5||^2\}$ to 20 equally spaced points between 0 and 10.

given by the Cholesky decomposition will look quite different. A particular efficient ordering is obtained by using a Cholesky factorization with pivoting (Dongarra *et al.*, 1978). The resulting basis vectors are shown in Figure 4. In this example, nearly all of the variation in z is explained by the first 5 basis vectors.

A third (of many more) alternatives can be obtained by taking the singular value decomposition (SVD) of Σ for which $\Sigma = UDU^T$ where U is an orthonormal matrix and D is diagonal with non-negative elements. The square root matrix can then be given by $L = UD^{\frac{1}{2}}$. The columns of the resulting matrix L are shown in Figure 5. As with the Cholesky factorization with pivoting, this approach gives an efficient basis representation of z. Sections 4 and 5 consider modeling approaches that rely on efficient basis representations of z.

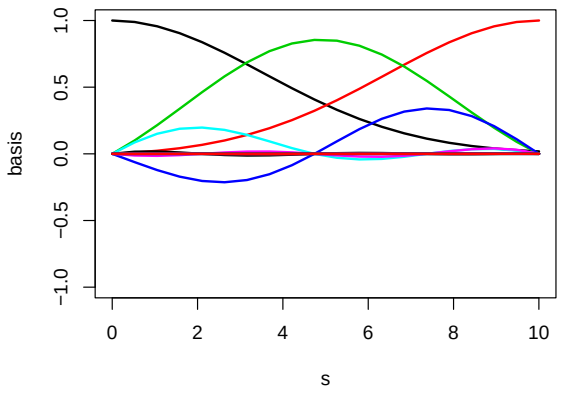


Figure 4: Basis vectors resulting from the Cholesky factorization with pivoting of the covariance matrix obtained by applying the covariance rule $\Sigma_{ij} = \exp\{-||(s_i - s_j)/5||^2\}$ to 20 equally spaced points between 0 and 10.

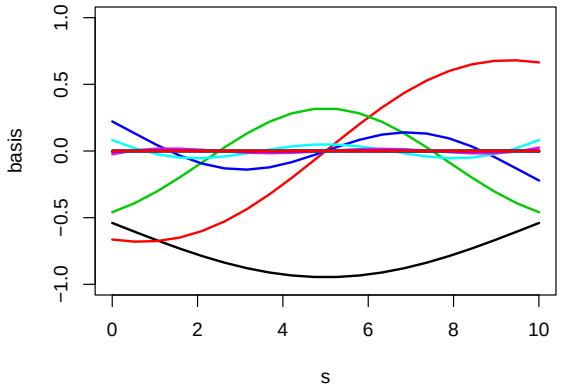


Figure 5: Basis vectors resulting from the SVD factorization with pivoting of the covariance matrix obtained by applying the covariance rule $\Sigma_{ij} = \exp\{-||(s_i - s_j)/5||^2\}$ to 20 equally spaced points between 0 and 10.

2.2 Conditional distributions

In nearly any spatial application, some aspect of the problem involves inferring about $z(s)$, over a spatial domain $\mathcal{S}$. To be concrete, suppose $z(s),\ s \in \mathcal{S}$ is a mean 0 GP with covariance rule $\Sigma_{ij} = \exp\{-||(s_i - s_j)||^2\}$, and $\mathcal{S}$ denotes $m = 8$ spatial locations $\{s_1, \ldots, s_m\} = \{0, 1, \ldots, 7\}$. Now we observe $z(s_i)$ at $n = 2$ spatial locations $s_2 = 1$ and $s_5 = 4$ as shown in Figure 6. Now, $z(s_1)$ and $z(s_5)$ are known. It is the conditional

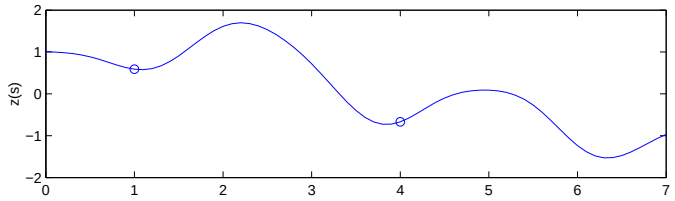


Figure 6: From the Gaussian process, we observe the values $z(s_2)$ and $z(s_5)$ and now want to determine the conditional distribution of the entire process $z(s)$.

distribution of the remaining sites in $\mathcal{S}$ that is required.

Standard normal computations will give us the conditional distribution of $z(s),\ s \in \{\mathcal{S} \setminus \{s_1, s_5\}\}$. After reordering, we have

$$\begin{pmatrix} z(s_2) \\ z(s_5) \\ \hline z(s_1) \\ z(s_3) \\ z(s_4) \\ z(s_6) \\ z(s_7) \\ z(s_8) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \left(\begin{array}{cc|ccc} 1 & .0001 & .3679 & \cdots & 0 \\ .0001 & 1 & 0 & \cdots & .0001 \\ \hline .3679 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \vdots & \ddots & \vdots \\ 0 & .0001 & 0 & \cdots & 1 \end{array}\right) \right).$$

Generally, for a mean 0 normally distributed vector, if

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

then

$$z_2|z_1 \sim N(\Sigma_{21}\Sigma_{11}^{-1}z_1, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}). \qquad (1)$$

See Anderson (1984), Section 2.5 for the derrivation. So if we take the first two components of the 8-vector to be z_1, and the remaining six to be z_2, we have the conditional mean of $z_2|z_1 = \mu_{2|1}$ given by $\Sigma_{21}\Sigma_{11}^{-1}z_1$ and the conditional variance of $z_2|z_1 = \Sigma_{2|1}$ given by $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

Instead of specifying $\mathcal{S}$ to be 8 points equally spaced between 0 and 7, we could define $\mathcal{S}$ to be a large number of points equally spaced between 0 and 7. Equation (1) above just as easily gives the conditional mean and variance of $z(s)$ over this $\mathcal{S}$. Of course, drawing from (1) may become computationally demanding if the number of prediction sites is very large. Figure 7 shows the conditional mean of $z(s)$ after

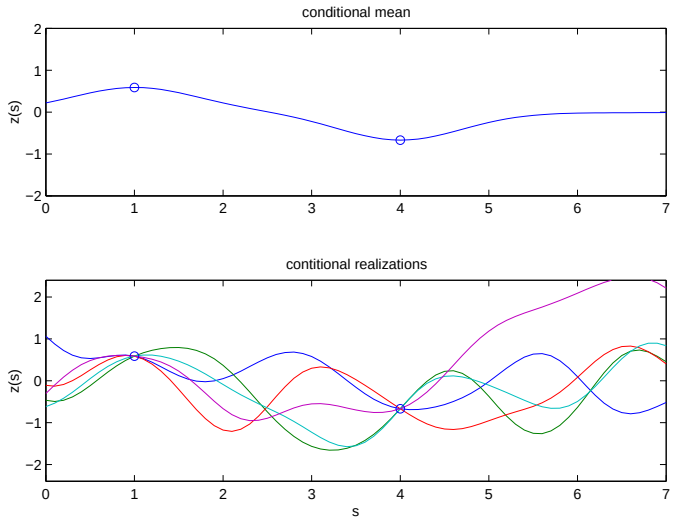


Figure 7: The conditional mean of $z(s)$ after observing $z(s)$ at the two spatial locations $s = 1$ and $s = 4$.

conditioning on the two points $z(s = 1)$ and $z(s = 4)$. The

figure also shows 4 draws from the GP $z(s)$ conditional on these two observations.

Equation (1) is one of the foundations for working with spatial models. Figures 8 and 9 show draws from the conditional process over the one-dimensional space $\mathcal{S} = [0, 15]$

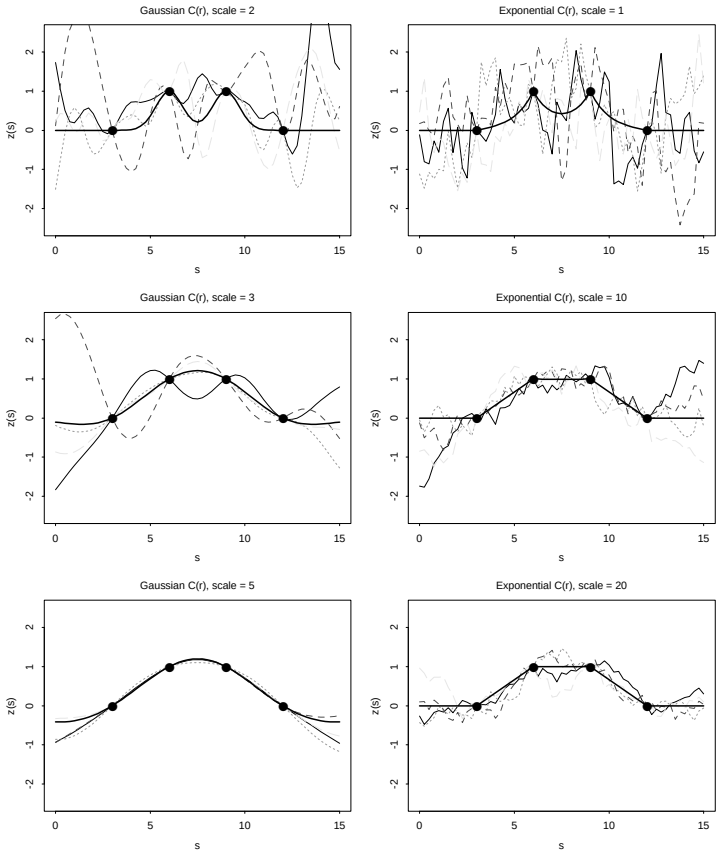


Figure 8: Conditional realizations of $z(s)$ under different covariance models after observing the four datapoints given by the black dots. The conditional mean of $z(s)$ is given by the thick black line; four conditional realizations are given by the dashed lines.

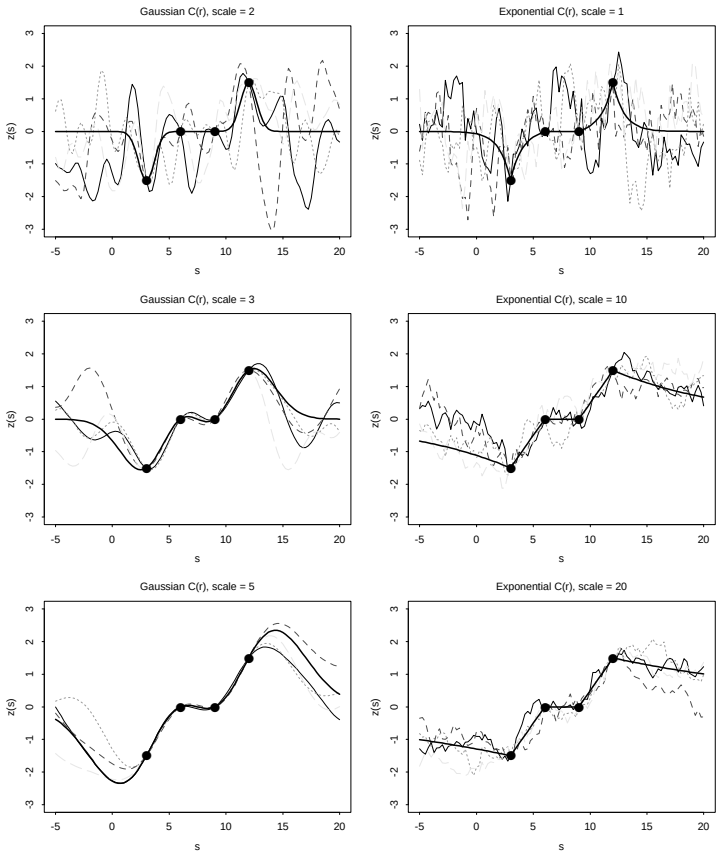


Figure 9: Conditional realizations of $z(s)$ under different covariance models after observing the four datapoints given by the black dots. The conditional mean of $z(s)$ is given by the thick black line; four conditional realizations are given by the dashed lines.

after conditioning on the same 4 points. In each plot, a different covariance rule is used. The table below shows the covariance rule used for each plot in Figures 8 and 9.

covariance rules/functions for Figs 8 and 9

left plots	right plots
$\Sigma_{ij} = \exp\{-\lvert(s_i - s_j)/2\rvert^2\}$	$\Sigma_{ij} = \exp\{-\lvert(s_i - s_j)/1\rvert^1\}$
$\Sigma_{ij} = \exp\{-\lvert(s_i - s_j)/3\rvert^2\}$	$\Sigma_{ij} = \exp\{-\lvert(s_i - s_j)/10\rvert^1\}$
$\Sigma_{ij} = \exp\{-\lvert(s_i - s_j)/5\rvert^2\}$	$\Sigma_{ij} = \exp\{-\lvert(s_i - s_j)/20\rvert^1\}$

Note the left-hand column of plots correspond to GP's with a *Gaussian* covariance function; the right-hand column of plots correspond to GP's with an *exponential* covariance function. Realizations from this exponential process are rougher as compared to the realizations from GP's with the Gaussian covariance rule.

Equation (1) is just as applicable to higher dimensional GP's. As an example, the top left plot in Figure 10 shows a realization from a GP over the 2-d lattice $\mathcal{S} = \{1, 2, \ldots, 20\}^2$ with covariance given by $\Sigma_{ij} = \exp\{-|d_{ij}/2|^2\}$. Here d_{ij} is the Euclidean distance between points $s_i = (s_{i1}, s_{i2})$ and $s_j = (s_{j1}, s_{j2})$ in $\mathcal{S}$. If we condition on the 20 points corresponding to $s_2 = 1$, the top right plot shows the conditional mean of $z(s)$ and the bottom two figures each show a realization of the process conditional on the 20 edge values $z(s_1, s_2 = 1)$.

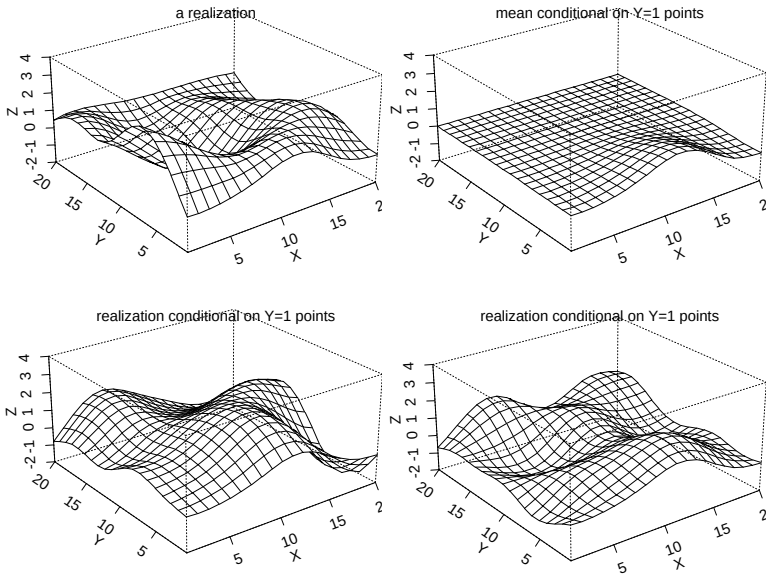


Figure 10: Top left: A realization from a 2-d Gaussian process with mean zero and covariance function $C(r) = \exp\{-(r/2)^2\}$. Top right: The conditional mean of the process after conditioning on the points for which $y = 1$. Bottom row: Realizations of the Gaussian process after conditioning on the edge points at $y = 1$.

2.3 Soft conditioning

In almost any actual application, the underlying spatial process $z(s)$ is never measured with complete accuracy at any spatial location. Hence we need to be able to characterize our uncertainty about $z(s)$ given noisy observations. This chapter approaches the problem from a Bayesian perspective.

For this example, we consider the mean 0 GP $z(s)$ over the 8 locations $\mathcal{S} = \{0, 1, \ldots, 7\}$ with covariance rule $\Sigma_{ij} = \exp\{-|s_i - s_j|^2\}$. We take the observed data y now to be a noisy version of z over each point in $\mathcal{S}$

$$y(s_i) = z(s_i) + \epsilon(s_i),\ i = 1, \ldots, n,$$

with

$$\epsilon(s_i) \overset{iid}{\sim} N(0, \sigma_y^2),\ i = 1, \ldots, n.$$

The data are shown in Figure 11 with error bars that correspond to $\pm 2\sigma_y$. Also shown is the true underlying process

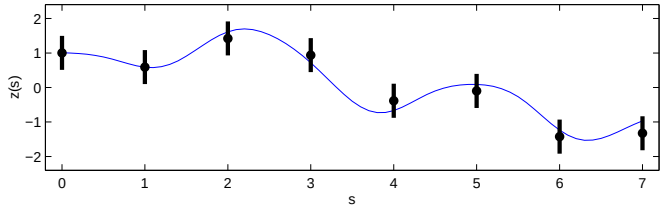


Figure 11: Data from a noisy observation of the continuous Gaussian process $z(s)$ at locations $s_i = i,\ i = 0, \ldots, 7$. The dots denote the observations and the bars show $\pm 2\sigma_y$.

$z(s)$ about which we wish to infer.

Under the Bayesian paradigm, inference about $z(s)$ is based on its posterior distribution given the $n = 8$ noisy observations y. The likelihood – or sampling model – for the observed data is

$$L(y|z) \propto |\Sigma_y|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y - z)^T \Sigma_y^{-1}(y - z)\}$$

where $\Sigma_y = \sigma_y^2 I_n$ and I_n denotes the $n \times n$ identity matrix. The GP prior model for $z(s)$ is

$$\pi(z) \propto |\Sigma_z|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2} z^T \Sigma_z^{-1} z\}$$

where Σ_z is obtained by applying the covariance rule to the $n = 8$ spatial locations in $\mathcal{S}$. The resulting posterior density for z given y is proportional to the product of the Likelihood and prior

$$\begin{aligned} \pi(z|y) &\propto L(y|z) \times \pi(z) \qquad (2) \\ &\propto \exp\{-\tfrac{1}{2} z^T (\Sigma_y^{-1} + \Sigma_z^{-1}) z + z^T \Sigma_y^{-1} y + f(y)\}. \end{aligned}$$

The unnormalized density above is that of a normal

$$z|y \sim N(V\Sigma_y^{-1} y, V)$$

where $V = (\Sigma_y^{-1} + \Sigma_z^{-1})^{-1}$. Note the posterior mean

$$V\Sigma_y^{-1} y = (\Sigma_y^{-1} + \Sigma_z^{-1})^{-1} \Sigma_y^{-1} y$$

is a precision weighted average of the prior mean (0) and the data. The identity

$$(\Sigma_y^{-1} + \Sigma_z^{-1})^{-1} \Sigma_y^{-1} = \Sigma_z (\Sigma_y + \Sigma_z)^{-1}$$

can be helpful for computations.

Often, the noisy observations at $\{s_1, \ldots, s_n\}$ will be used to obtain the posterior distribution for $z(s)$ at a collection of m unobserved sights $\{s_1^*, \ldots, s_m^*\}$. The previous recipe for obtaining $\pi(z|y)$ can easily be modified to accommodate this situation.

We define

$$\begin{aligned} y^d &= (y(s_1), \ldots, y(s_n))^T \\ z^d &= (z(s_1), \ldots, z(s_n))^T \\ y^* &= (y(s_1^*), \ldots, y(s_m^*))^T \\ z^* &= (z(s_1^*), \ldots, z(s_m^*))^T \\ y &= (y^d; y^*) \\ z &= (z^d; z^*) \end{aligned}$$

so that y holds a $n+m$ vector of "observations" (the last m components of y are only notional), and z holds $z(s)$ restricted to the observation and prediction locations.

The sampling model for y is the same as before

$$L(y|z) \propto |\Sigma_y|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y-z)^T \Sigma_y^- (y-z)\}$$

except now

$$\Sigma_y^- = \begin{pmatrix} \frac{1}{\sigma_y^2} I_n & 0 \\ 0 & 0_{m\times m} \end{pmatrix}$$

and $|\Sigma_y|$ is defined to be σ^{2n}. Here zero precisions (infinite variances) are specified for the components of y which are unobserved. Also, the GP prior is exactly as before for z,

$$\pi(z) \propto |\Sigma_z|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2} z^T \Sigma_z^{-1} z\}$$

but now Σ_z is the $(n+m)\times(n+m)$ covariance matrix obtained by applying the covariance rule to the augmented set of spatial locations–both observation and prediction locations.

The resulting posterior distribution for $z = (z^d, z^*)$ is then

$$z|y \sim N(V\Sigma_y^- y, V)$$

where

$$V = (\Sigma_y^- + \Sigma_z^{-1})^{-1}.$$

Note the posterior distribution for z only depends on Σ_y through its inverse which is well defined. Figure 12 shows realizations from $\pi(z|y)$ where the prediction sites are a fine grid of locations between 0 and 7.

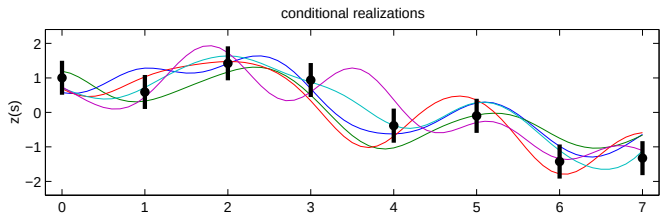


Figure 12: Conditional realizations of $z(s)$ after the noisy observations given in Figure 11.

2.3.1 An example

As an example we have $n = 120$ spatially located log-dioxin concentration measurements taken from a 100m $\times$ 200m contaminated site in Missouri, USA (Ryti *et al.*, 1993) shown in the left plot of Figure 13.

We wish to predict the concentrations over a 15×30 grid denoted by the points in left plot of the figure. We specify a mean 0 GP prior for $z(s)$ – the log-concentration surface with covariance rule

$$\Sigma_{ij} = \exp\{-||(s_i - s_j)/15||^2\}.$$

The posterior mean for the $m = 15 \cdot 30$ prediction sites is shown in the center plot. The prediction standard deviations, given by the square root of the diagonal of the last $m \times m$ submatrix of the posterior variance matrix V, is shown in the right plot.

Here we specified the prior mean of $z(s)$ as well as the form of the covariance rule. An important component of spatial modeling – ignored here, for now – is determining these

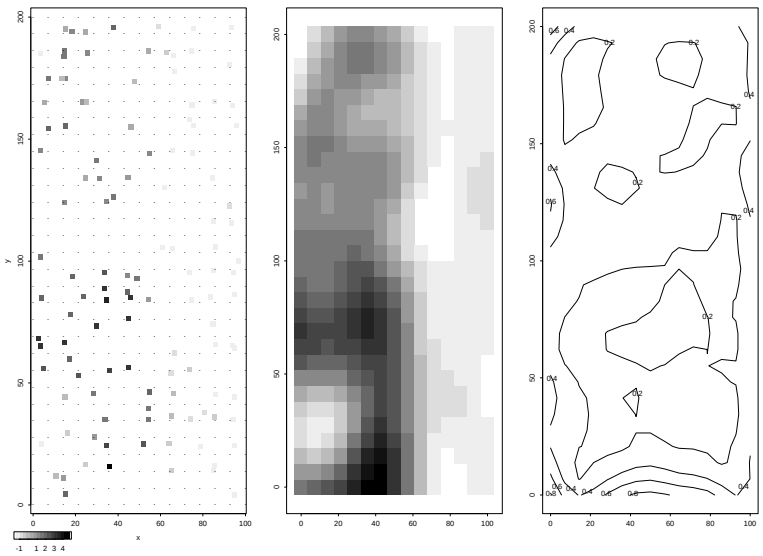


Figure 13: Left: log dioxin concentration measurements from the Piazza Road pilot study. Middle: posterior mean for the concentration at the grid sites denoted by dots in the left hand figure. Right: pointwise posterior standard deviation of $z(s)$ given the measurements.

features of the spatial model. The geostatistical references mentioned in Section 1 have much more to say on this topic.

3 Gaussian Markov random fields and Bayesian computation

The posterior distribution of (2) depends on precision (inverse covariance) matrices in the prior and likelihood, and can be trivially rewritten

$$\pi(z|y) \quad \propto \quad \exp\{-\frac{1}{2}z^T(W_y + W_z)z + z^T W_y y + f(y)\}$$

where $W_y = \Sigma_y^{-1}$ and $W_z = \Sigma_z^{-1}$. Hence

$$z|y \sim N(\Lambda^{-1}W_y y, \Lambda^{-1})$$

where $\Lambda = W_y + W_z$.

Rather than specifying a the prior dependence structure for the spatial process $z(s)$ through a covariance rule, one can specify the dependence through the precision matrix W_z. As we'll see in this section, the is a natural connection between W_z and the conditional dependence structure of $z(s)$. These *Gaussian Markov random field* (GMRF) specifications can be very effective for large systems, such as image models, because they are amenable to efficient computational approaches. The strong links between GMRF's and MCMC make it natural to introduce Gibbs sampling in this section as well.

A GMRF is defined over a fixed set of sites $s_1, \ldots, s_n$. The vector $z = (z(s_1), \ldots, z(s_n))^T$ is the spatial process to be modeled with a GMRF specification. We now list some notation and facts regarding a GMRF specification for $z(s)$ – a far more detailed description of GMRF's can be found in Rue and Held (2005):

- z_i is the value of $z(s)$ at site s_i;
- z_{-i} is the $n-1$-vector $(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n)^T$;
- $i \sim j$ means s_i and s_j are neighbors;

- $z_{\partial i} = \{z_j : \ i \sim j\}$ (i.e.. all z_j for which s_j and s_i are neighbors);
- $\pi(z_i|z_{-i})$ is called the *full conditional* distribution of z_i (ie. its distribution given all other components of z);
- $\pi(z_i|z_{-i}) = \pi(z_i|z_{\partial i})$ ie. the full conditional for z_i depends only on its neighbors.

A Markov random field (MRF) is defined by specifying a neighborhood system and a set of n conditional densities $\{\pi(z_i|z_{\partial i}),\ i = 1, \ldots, n\}$. Figure 14 gives examples of neighborhood specifications on regular lattices. Note that not just

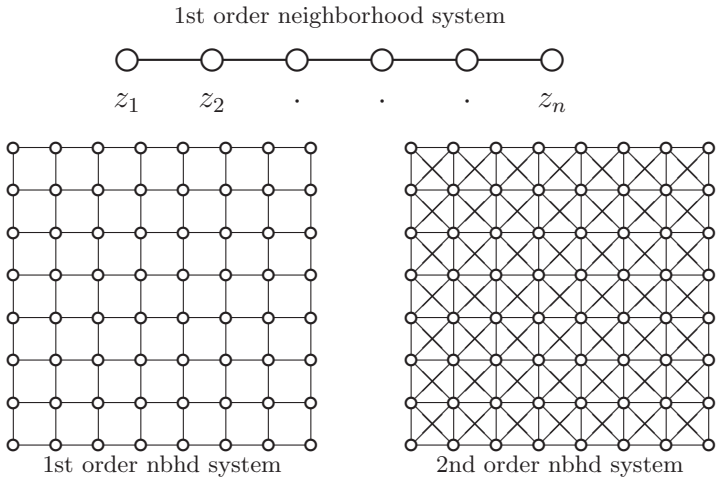


Figure 14: Top: A 1st order (nearest neighbor) neighborhood system for a regular 1-d lattice. Bottom left: A 1st order neighborhood system for a regular 2-d lattice. Bottom right: A 2nd order (nearest two neighbors) neighborhood system for a regular 2-d lattice.

any set of n full conditionals will yield a valid joint distribution $\pi(z)$. This section uses only Gaussian MRF's whose

full conditionals are determined by the precision matrix W_z. Section 6 gives an example of a non-Gaussian MRF.

3.1 Locally linear Gaussian MRF's

For a given neighborhood system define

$$z_i | z_{\partial i} \sim N(\bar{z}_{\partial i}, 1/n_i)$$

where n_i is the number of neighbors belonging to z_i and $\bar{z}_{\partial i}$ is the average of the z_js neighboring z_i. The full conditional density is then

$$\pi(z_i | z_{\partial i}) \propto \exp\left\{ -\frac{1}{2} \sum_{j \in \partial i} (z_i - z_j)^2 \right\}$$

which implies a joint Gaussian distribution for z:

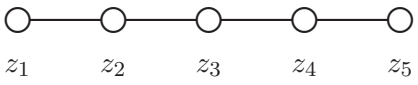
$$\begin{aligned} \pi(z) &\propto |W_z|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} z^T W_z z \right\} \qquad (3) \\ &\propto |W_z|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} \sum_{i \sim j} (z_i - z_j)^2 \right\} \end{aligned}$$

where the ij component of W_z is given by

$$W_{zij} = \begin{cases} n_i & \text{if } i = j \\ -1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

and the sum $\sum_{i \sim j} (z_i - z_j)^2$ is over each edge pair in the neighborhood system. This locally linear MRF model for $z(s)$ gives an alternative to the covariance based models described in the previous section. Again, any practical application of a MRF model will require that $\pi(x)$ include a parameter to scale $z(s)$. This is addressed addressed at the end of Section 3.2 and beyond.

As an example, if we take z to be the $n = 5$-site, 1st order system below



then the precision matrix W_z has the form

$$W_z = \begin{pmatrix} 1 & -1 & \circ & \circ & \circ \\ -1 & 2 & -1 & \circ & \circ \\ \circ & -1 & 2 & -1 & \circ \\ \circ & \circ & -1 & 2 & -1 \\ \circ & \circ & \circ & -1 & 1 \end{pmatrix}.$$

As a 1-d example, suppose there is an underlying true process $z = (z_1, z_2, \ldots, z_{25})^T$ at spatial sites $\{1, 2, \ldots, 25\}$ and we have noisy observations $y = (y_1, y_2, \ldots, y_{25})^T$ at each site. This gives the sampling model

$$L(y|z) \propto |W_y|^{\frac{1}{2}} \exp\{-\frac{1}{2}(y - z)^T W_y (y - z)\}$$

where $W_y = \frac{1}{4}I$, corresponding to noise variance of 4. We specify a first order, locally linear GMRF for z so that the prior has the form

$$\pi(z) \propto |W_z|^{\frac{1}{2}} \exp\{-\frac{1}{2} z^T W_z z\}.$$

Hence, the resulting posterior distribution for z given y is multivariate Gaussian

$$z|y \sim N(\Lambda^{-1} W_y y, \Lambda^{-1}),$$

where $\Lambda = W_y + W_z$. Figure 15 shows the true underlying process z along with the noisy observations y in the top left frame. Also shown in the figure are the posterior mean of z, some realizations from the posterior, and pointwise 90% intervals for z at each site.

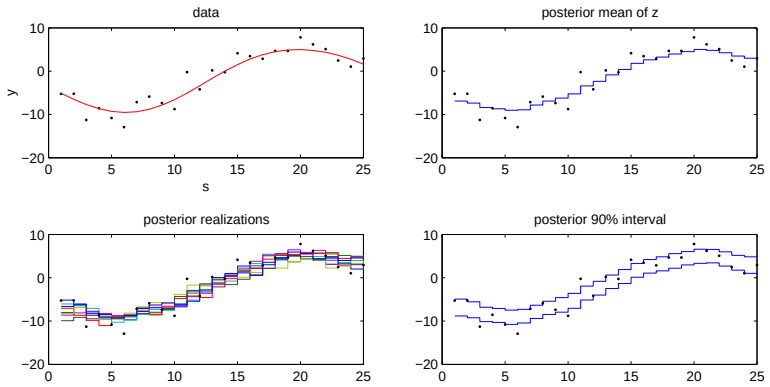
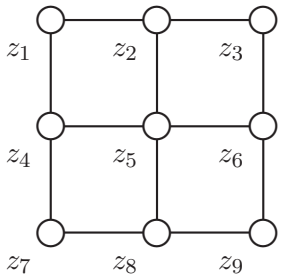


Figure 15: Data and posterior quantities for z which is given a first order, locally linear, Gaussian MRF prior (3). The data are independently scattered about the true process (top left plot) according to a $N(0, 4)$ distribution.

The locally linear specification for a 2-d lattice is also quite simple. For the $n = 9$ site 2-d lattice below



the locally linear specification leads to a 9×9 precision matrix

of the form

$$W_z = \begin{pmatrix} 2 & -1 & \circ & -1 & \circ & \circ & \circ & \circ & \circ \\ -1 & 3 & -1 & \circ & -1 & \circ & \circ & \circ & \circ \\ \circ & -1 & 2 & \circ & \circ & -1 & \circ & \circ & \circ \\ -1 & \circ & \circ & 3 & -1 & \circ & -1 & \circ & \circ \\ \circ & -1 & \circ & -1 & 4 & -1 & \circ & -1 & \circ \\ \circ & \circ & -1 & \circ & -1 & 3 & \circ & \circ & -1 \\ \circ & \circ & \circ & -1 & \circ & \circ & 2 & -1 & \circ \\ \circ & \circ & \circ & \circ & -1 & \circ & -1 & 3 & -1 \\ \circ & \circ & \circ & \circ & \circ & -1 & \circ & -1 & 2 \end{pmatrix}.$$

A 2-d example is shown in Figure 16 below. We generate data y over a 20×20 lattice which is the true process z (shown in the left plot of Figure 16) plus independent $N(0, 4)$ errors.

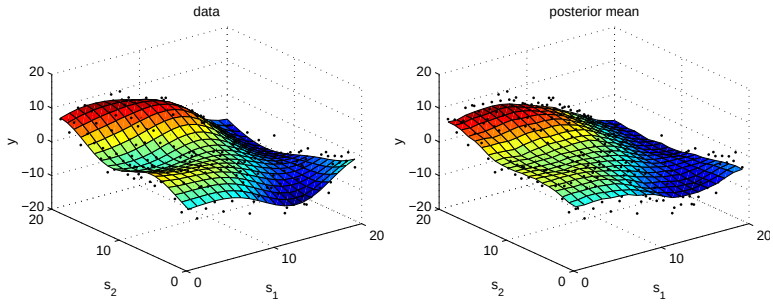


Figure 16: Left: Data and and true underlying process z over a 20×20 lattice. Right: Posterior mean for z resulting from the 2-d, first order, locally linear, Gaussian MRF prior on z (3). As with Figure 15, the data are independently scattered about the true process (left plot) according to a $N(0, 4)$ distribution.

3.2 General Gaussian MRF's

The locally linear GMRF specification is applicable in a wide range of applications. However the details of a particular application may require more structure than this prior can deliver. A more general form for a GMRF is

$$\pi(z) \propto |W|^{\frac{1}{2}} \exp\left\{-\frac{1}{2} z^T W z\right\}$$

where W is any symmetric, positive (semi) definite matrix. This model has full conditional distributions

$$z_i | z_{-i} \sim N\left(-\frac{1}{W_{ii}} \sum_{j \neq i}^{n} W_{ij} z_j, \frac{1}{W_{ii}}\right).$$

Specifying W determines the neighborhood structure of the system – $W_{ij} \neq 0 \quad \Leftrightarrow \quad$ sites i and j are neighbors. For details regarding how one might construct alternative precision matrices see Besag *et al.*, (1995), Besag and Kooperberg (1995), and Rue and Tjelmeland (2002).

It's worth pointing out that the locally linear GMRF specification results in an *intrinsic* prior distribution for z – that is, the model is not proper since $\pi(z) = \pi(z + c\mathbf{1})$ for any constant c (of course, the resulting posterior is proper). A proper alternative to the locally linear GMRF is one for which the locally linear precision W_z is replaced by $W_z + \delta I$. Figure 17 shows realizations from the stationary ($\delta = 0.1$) and intrinsic versions of the locally linear prior. These realizations are scaled so that the increments between adjacent sizes are similar. Even though the 2 realizations look similar if one only compares increment magnitude, their global nature is clearly quite different.

As an example of a possible GMRF specification, we take the locally quadratic prior. Here the elements of W are chosen so that $E(z_i | z_{\partial i})$ is the best quadratic fit at s_i given

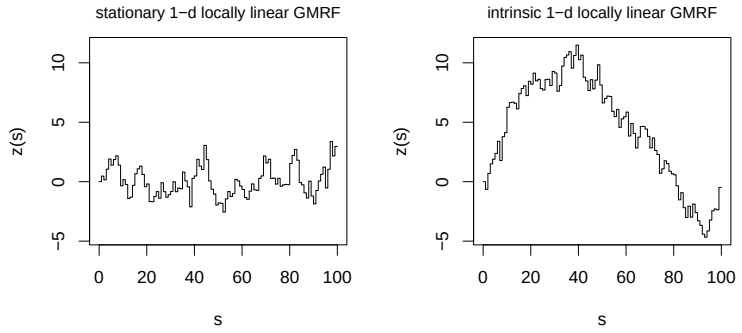


Figure 17: Realizations from a stationary (left) and intrinsic (right) locally linear GMRF priors. The stationary prior has mean 0 and precision proportional to $W + 0.1 \cdot I$. The draws are conditional on $z_1 = 0$. The stationary realization is scaled so that the expected squared increment size $E(z_i - z_{i+1})^2$ is the same in both plots.

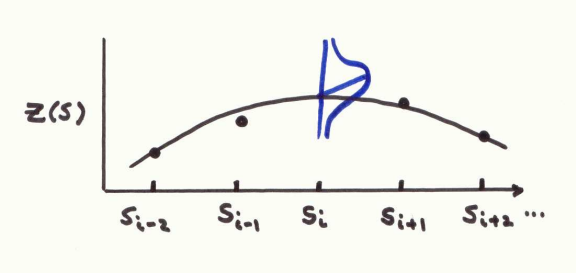


Figure 18: The locally quadratic GMRF specification defines the precision matrix W so that $E(z_i|z_{\partial i})$ is the fit of the parabola estimated from the four pairs (s_j, z_j), $j = i-2, i-1, i+1, i+2$.

$z_{\partial i}$ (see Figure 18). The four pairs of points (s_{i-2}, z_{i-2}), (s_{i-1}, z_{i-1}), (s_{i+1}, z_{i+1}) and (s_{i+2}, z_{i+2}) are used to estimate a parabola of the form $\hat{z}(s) = b_1 s + b_2 s^2$. The fit at site s_i is just a linear combination of the 4 neighboring z_j's. The re-

sulting precision matrix is then band diagonal with 6's along the diagonal, -4's on the one-off diagonals, and 1's on the two off diagonals. The precision matrix needs to be modified to account for edge sites. Note that the locally linear GMRF specification defines W so that $E(z_i|z_{\partial i})$ is the fit of the line estimated from the two pairs (s_j, z_j), $j = i-1, i+1$. Later in this section some comparisons between posteriors obtained under the locally quadratic and locally linear prior formulations are made.

Finally, up to now, the GMRF priors depend only on the precision matrix W_z. In any real application, it is useful to allow for a parameter λ_z to scale this precision matrix. From now on, a GMRF prior for z will typically include the precision scalar so that

$$\pi(z|\lambda_z) \propto \lambda_z^{\frac{n}{2}} \exp\{-\frac{1}{2}\lambda_z z^T W_z z\}. \tag{4}$$

The previous examples were fortuitously scaled so that $\lambda_z = 1$ was a good choice. The precision parameter λ_z controls the regularity of the process $z(s)$. Note the exponent in the term $\lambda_z^{\frac{n}{2}}$ in (4) is used here. More generally, it may be preferable to use rank(W) rather than n since W is often not of full rank. This distinction makes little difference when n is large; we use n throughout this primer.

3.3 Computing with large systems

Recall, in each of the examples in the previous section the resulting posterior was of the form:

$$z|y \sim N(\Lambda^{-1}W_y y, \Lambda^{-1}) \tag{5}$$

where $\Lambda = W_y + \lambda_z W_z$ and λ_z scales the spatial process z. In cases where n is quite large, computing the posterior mean $\Lambda^{-1}W_y y$ can be prohibitive. Here the local nature of the MRF specification can be quite helpful.

The full conditionals of the density implied by (5) are

$$\pi(z_i|z_{\partial i}, y) \propto \exp\left\{ -\frac{1}{2}\lambda_y(y_i - z_i)^2 - \frac{1}{2}\lambda_z W_{ii} z_i^2 - \lambda_z \sum_{j\in\partial i} W_{ij} z_i z_j \right\}$$

which implies

$$z_i|z_{\partial i}, y \sim N\left(\frac{\lambda_y y_i + \lambda_z \sum_{j\in\partial i} W_{ij} z_j}{\lambda_y + \lambda_z W_{ii}}, \frac{1}{\lambda_y + \lambda_z W_{ii}} \right).$$

In the case of the locally linear GMRF, this simplifies to

$$z_i|z_{\partial i}, y \sim N\left(\frac{\lambda_y y_i + n_i \lambda_z \bar{z}_{\partial i}}{\lambda_y + n_i \lambda_z}, \frac{1}{\lambda_y + n_i \lambda_z} \right)$$

where

$$\begin{aligned} n_i &= \#\text{ of neighbors belonging to } s_i \text{ and} \\ \bar{z}_{\partial i} &= \frac{1}{n_i} \sum_{j\in\partial i} z_j. \end{aligned}$$

3.3.1 Iterated conditional modes

The method of iterated conditional modes (ICM) (Besag, 1986) uses the univariate full conditional distributions in an iterative scheme to determine the mode of the posterior distribution $\pi(z|y)$. The algorithm iteratively replaces each z_i by the mode of it's full conditional:

for $t = 1, \ldots,$ nscan {
    for $i = 1, \ldots, n$ {

$$\text{set } z_i = \frac{\lambda_y y_i + n_i \lambda_z \bar{z}_{\partial i}}{\lambda_y + n_i \lambda_z}$$

    }
}.

Here the newly updated z_js should be used when computing $\bar{z}_{\partial i}$.

In general, the eventual value $\hat{z}_{ICM}$ will correspond to a local mode of the posterior density $\pi(z|y)$. If $\pi(z|y)$ is Gaussian (as it is here) $\hat{z}_{ICM}$ will correspond to the posterior mean, which is the global optimum of $\pi(z|y)$. Hence, ICM iteratively solves the posterior mean equation $\hat{z}_{ICM} = \Lambda^{-1} W_y y$.

Figure 19 shows ICM being carried out on a 1-d example. The spatial process z is initialized at the observed data y. By the 10th scan through the vector z, the posterior mean solution is reached.

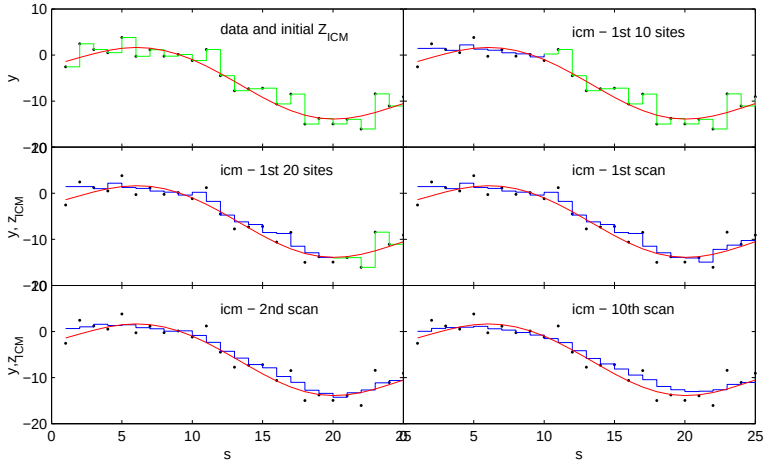


Figure 19: ICM applied to a 1-d example where $y = z + \epsilon$ and z is given a locally linear GMRF prior so that the resulting posterior distribution is given by (5). By the 10th scan through the data, the ICM solution is reached.

3.3.2 Gibbs sampling

While ICM allows one to compute the posterior mean, Markov chain Monte Carlo (MCMC) gives an approach for generating realizations from the posterior distribution. With realiza-

tions, one can explore many features of $\pi(z|y)$ – in particular, uncertainties regarding z can be assessed.

The *Gibbs sampler* is a MCMC scheme which iteratively replaces each z_i by a draw from its full conditional:

for $t = 1, \ldots,$ nscan {
  for $i = 1, \ldots, n$ {

$$\text{draw } z_i \sim N\left(\frac{\lambda_y y_i + n_i \lambda_z \bar{z}_{\partial i}}{\lambda_y + n_i \lambda_z}, \frac{1}{\lambda_y + n_i \lambda_z}\right)$$

  }
}

Figure 20 shows a 1-d example of sampling from the posterior resulting from a locally linear GMRF prior for z over 25 regularly spaced sites. The unknown spatial process z is initialized at 0. The MCMC continues, giving a dependent sequence of T realizations $z^1, \ldots, z^T$ which can be treated as draws from the posterior distribution $\pi(z|y)$. Of course, the early realizations will be influenced by the initialization of z. Hence it is sensible to discard these early MCMC draws until the sequence has "forgotten" the initialization value of z.

Accounting for unknown model parameters

In most applications, there are various nuisance parameters that need to be incorporated into the model. In the examples of this section, such parameters might include the precision of the observation measurements and the scaling of the GMRF prior for $z(s)$.

Assume the data y are formed corresponding to the model

$$y = z + \epsilon, \text{ where } \epsilon \sim N(0, \frac{1}{\lambda_y} I_n).$$

Hence, the likelihood

$$L(y|z, \lambda_y) \propto \lambda_y^{\frac{n}{2}} \exp\{-\tfrac{1}{2}\lambda_y (y - z)^T (y - z)\}$$

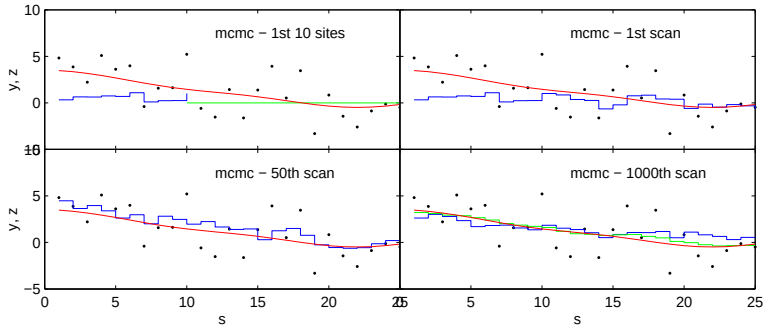


Figure 20: Gibbs sampling applied to the posterior distribution (5) resulting from the data y (denoted by the dots) and a locally linear GMRF specification for the underlying spatial process z. The data are generated from the model $y = z + \epsilon$, where the true value for z is given by the smooth red line. z is initialized at 0. The green line in the bottom right plot shows the posterior mean estimated from the 1000 MCMC realizations.

has the precision parameter λ_y that describes the size of the observation errors. Similarly, the GMRF prior for z has the form

$$\pi(z|\lambda_z) \propto \lambda_z^{\frac{n}{2}} \exp\{-\tfrac{1}{2}\lambda_z z^T W_z z\}$$

where the precision λ_z controls the regularity in the spatial process z.

In order to account for uncertainty regarding these precision parameters, uninformative, conjugate priors are specified them:

$$\begin{aligned} \pi(\lambda_y) &\propto \lambda_y^{a_y - 1} \exp\{-b_y \lambda_y\}; \\ \pi(\lambda_z) &\propto \lambda_z^{a_z - 1} \exp\{-b_z \lambda_z\}. \end{aligned}$$

Typically, $\pi(\lambda_y)$ and $\pi(\lambda_z)$ are chosen to be rather uninformative. The choice of $a_y = a_z = 1$ and $b_y = b_z = .0001$ will usually work well. If λ is expected to be very large, then a

smaller value b will be necessary. Note that using a value of a_z very close to 0 places a lot of prior mass for λ_z near 0. This can lead to spurrious posterior results (Gelman, 2005).

The resulting posterior distribution is proportional to the product of the likelihood and prior

$$\begin{aligned}\pi(z,\lambda_y,\lambda_z|y) &\propto L(y|z,\lambda_y)\times\pi(z|\lambda_z)\times\pi(\lambda_y)\times\pi(\lambda_z) \quad (6)\\ &\propto \lambda_y^{\frac{n}{2}}\exp\{-\tfrac{1}{2}\lambda_y(y-z)^T(y-z)\}\times\\ &\quad\ \lambda_z^{\frac{n}{2}}\exp\{-\tfrac{1}{2}\lambda_z z^T W_z z\}\times\\ &\quad\ \lambda_y^{a_y-1}\exp\{-b_y\lambda_y\}\times\lambda_z^{a_z-1}\exp\{-b_z\lambda_z\}\end{aligned}$$

which describes uncertainty regarding (z,λ_y,λ_z).

The full conditional densities of $\pi(z,\lambda_y,\lambda_z|y)$ are then

$$\begin{aligned}\pi(z|\lambda_y,\lambda_z,y) &\propto \exp\{-\frac{1}{2}\lambda_y(y-z)^T(y-z)-\frac{1}{2}\lambda_z z^T W_z z\}\\ \pi(\lambda_y|z,\lambda_z,y) &\propto \lambda_y^{n/2+a_y-1}\exp\{-[b_y+.5(y-z)^T(y-z)]\lambda_y\}\\ \pi(\lambda_z|z,\lambda_y,y) &\propto \lambda_z^{n/2+a_z-1}\exp\{-(b_z+.5z^T W_z z)\lambda_z\}.\end{aligned}$$

Each of these full conditionals corresponds to a standard distribution – normal for z and Gamma for the precisions.

$$\begin{aligned}z|\lambda_y,\lambda_z,y &\sim N(\Lambda^{-1}\lambda_y y,\Lambda^{-1}),\ \text{ where } \Lambda=\lambda_y I_n+\lambda_z W_z\\ \lambda_y|z,\lambda_z,y &\sim \Gamma(n/2+a_y, b_y+.5(y-z)^T(y-z))\\ \lambda_z|z,\lambda_y,y &\sim \Gamma(n/2+a_z, b_z+.5z^T W_z z)\end{aligned}$$

For z, one can also determine the full conditionals corresponding to each component.

$$z_i|z_{-i},\lambda_y,\lambda_z,y \sim N\left(\frac{\lambda_y y_i+n_i\lambda_z\bar{z}_{\partial i}}{\lambda_y+n_i\lambda_z},\frac{1}{\lambda_y+n_i\lambda_z}\right),\ i=1,\ldots,n$$

The Gibbs sampler implementation of MCMC for $\pi(z,\lambda_y,\lambda_z|y)$ simply initializes (z,λ_y,λ_z) and then iteratively samples from

each of the full conditionals. For the results shown in Figure 21, z was initialized at $\frac{1}{2}y + \frac{1}{2}\bar{y}$, and the two precisions were set to draws from their full conditionals given the initial value for the vector z. The Gibbs sampler was run for $T = 2000$ scans through each of the parameters, and the first 500 realizations were discarded to remove the effect of the initialization.

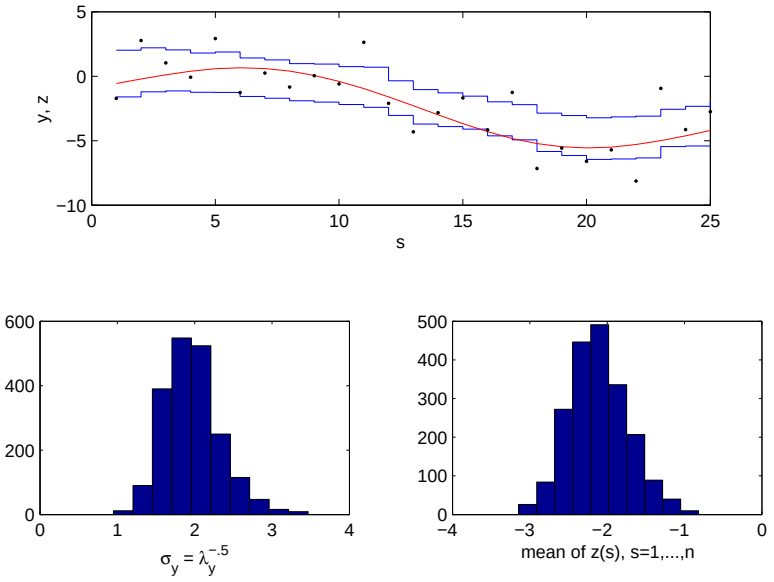


Figure 21: Summaries of the Gibbs sampling output produced from sampling the posterior distribution (6). Top: pointwise 90% credible intervals for z; Bottom left: a histogram of draws of the error standard deviation obtained by simply taking $\lambda_y^{-\frac{1}{2}}$ at each iteration; Bottom right: the posterior distribution of the average of z.

Figure 21 shows some summaries of the MCMC output produced by this Gibbs sampler: pointwise 90% credible intervals for each component of z; the posterior distribution for

$\sigma_y = \lambda_y^{-\frac{1}{2}}$; and the posterior distribution of $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$. Marginal distributions are simply estimated with histograms of the MCMC output. This example highlights how easy it is to infer about functions of z. Here we obtained the posterior distribution of the spatial mean of z. One could just as easily infer about other quantities such as the max of z, or the spatial location where the max of z is reached.

Figures 22 and 23 show posterior 80% credible intervals for z under two different formulations – the locally linear and the locally quadratic GMRFs. The spatial process resides on the integers from 1 to 50. Data were generated from the model $y = z + \epsilon$ where the true value of z is given by the solid line in the figures, and $\epsilon \sim N(0, \frac{1}{10} I_{50})$. The precision parameters λ_y and λ_z are treated as unknown and assigned $\Gamma(1, .0001)$ priors. Each row in the figures corresponds to a different realization for ϵ. The dotted lines connect the resulting data points. Generally, the locally linear formulation does a better job reconstructing z and more accurately estimates the error precision λ_y given the data in Figure 22. For the sine wave data in Figure 23 the locally quadratic prior leads to more faithful reconstructions.

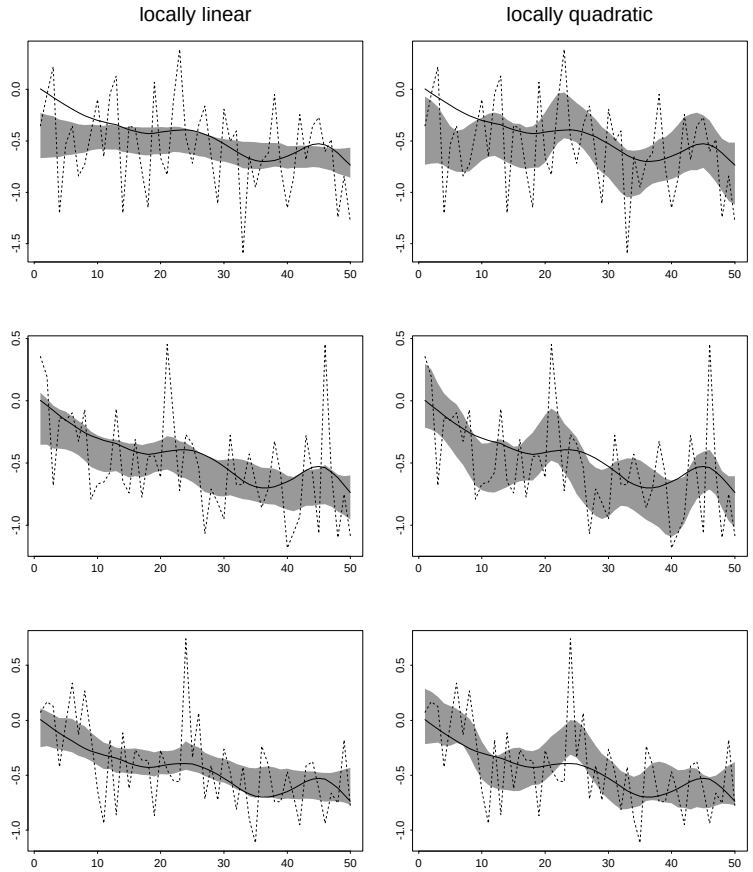


Figure 22: Posterior reconstructions under the locally linear and locally quadratic GMRF formulations over sites $\{1, 2, \ldots, 50\}$. In each row data (dotted lines) are generated according to the model $y = z + \epsilon$ where the true z is given by the solid line. Pointwise posterior 80% intervals are computed for z and shown by the gray regions in the plots. The left column shows credible intervals resulting from the locally linear formulation. The right column shows credible intervals resulting from the locally quadratic formulation.

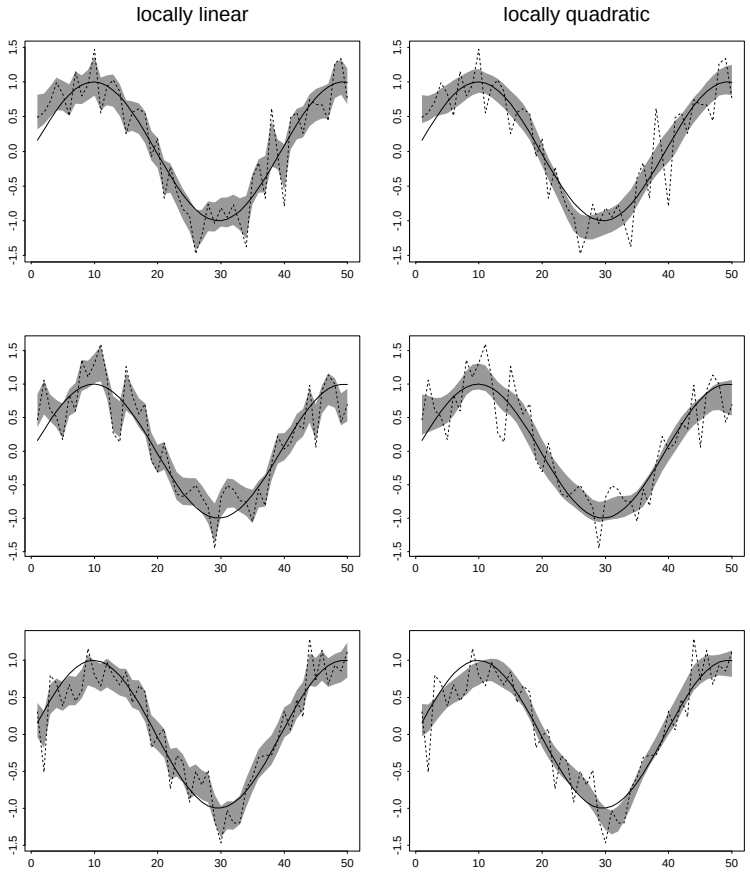


Figure 23: Posterior reconstructions under the locally linear and locally quadratic GMRF formulations over sites $\{1, 2, \ldots, 50\}$. In each row data (dotted lines) are generated according to the model $y = z + \epsilon$ where the true z is given by the solid line. Pointwise posterior 80% intervals are computed for z and shown by the gray regions in the plots. The left column shows credible intervals resulting from the locally linear formulation. The right column shows credible intervals resulting from the locally quadratic formulation.

4 Convolution-based spatial models

GMRF models work well for image and lattice data, however when data are irregularly spaced, a continuous model for the spatial process $z(s)$ is usually preferable. In this section convolution – or, equivalently, kernel – models are introduced. These models construct a continuous spatial model $z(s)$ by smoothing out a simple, regularly spaced latent process. In some cases, a GMRF model is used for this latent process.

The convolution process $z(s)$ is determined by specifying a latent process $x(s)$ and a smoothing kernel $k(s)$. We restrict the latent process $x(s)$ to be nonzero at the fixed spatial sites $\omega_1, \ldots, \omega_m$, also in $\mathcal{S}$ and define $x = (x_1, \ldots, x_m)^T$ where $x_j = x(\omega_j),\ j = 1, \ldots, m$. For now, the x_j's are modeled as independent draws from a $N(0, 1/\lambda_x)$ distribution. The resulting continuous Gaussian process is then

$$\begin{aligned} z(s) &= \int_{\mathcal{S}} k(u-s)dx(u) \\ &= \sum_{j=1}^{m} k(\omega_j - s)x_j \end{aligned} \tag{7}$$

where $k(\omega_j - \cdot)$ is a kernel centered at ω_j. This convolution model is depicted in Figure 24 below. Typically, $k(s)$ is taken

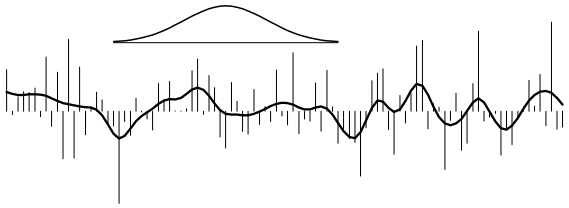


Figure 24: A one-dimensional Gaussian process obtained from smoothed white noise.

to be a normal kernel centered at 0, with a sd σ_k.

As long as $k(s)$ is symmetric, the same representation for $z(s)$ could be explained as a basis construction. Define the m basis functions $k_1(s), \ldots, k_m(s)$, where $k_j(s) = k(s - \omega_j)$ as shown in Figure 25. Here the $k_j(s)$'s are normal densities

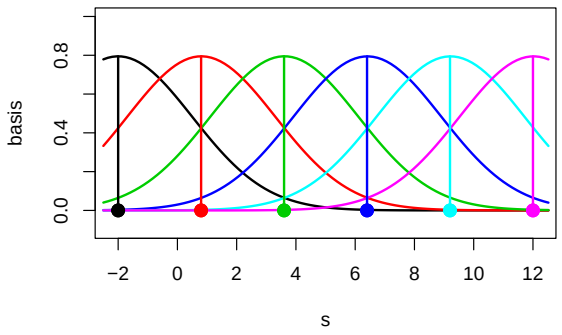


Figure 25: Basis functions to construct a 1-d spatial process. Here the $m = 6$ basis functions $k_j(s)$ are normal kernels centered at spatial locations ω_j given by the dots. The sd of the kernels is equal to the spacing of the ω_j's.

centered at spatial locations ω_j; the ω_j's are shown by the dots in Figure 25. The standard deviation of the normal density σ_k is set equal to the spacing between adjacent ω_j's. The continuous spatial process $z(s)$, $s \in \mathcal{S}$, with $\mathcal{S} = [0, 10]$ is defined as

$$z(s) = \sum_{j=1}^{m} k_j(s) x_j \text{ where } x \sim N(0, I_m). \tag{8}$$

Note the bsis representation in (8) is the foundation of a number of kernel and basis approaches for fitting data. See Hastie *et al.*(2001) for an overview of these methods.

If our interest is restricted to a finite set of n spatial locations $s_1, \ldots, s_n$, the n-vector $z = (z(s_1), \ldots, z(s_n))^T$ is given by

$$z = Kx \text{ where } K_{ij} = k_j(s_i). \tag{9}$$

Frequently, applications require the value of $z(s)$ over a grid of spatial locations. Hence, the discrete representation above can be quite useful.

With either interpretation, convolution or basis, the spatial process $z(s)$ is determined by the distribution of x, the latent sites $\omega_1, \ldots, \omega_m$, and the kernel $k(s)$. Figure 26 shows how $z(s)$ is constructed from a random draw of the $m = 6$-vector x according to (7).

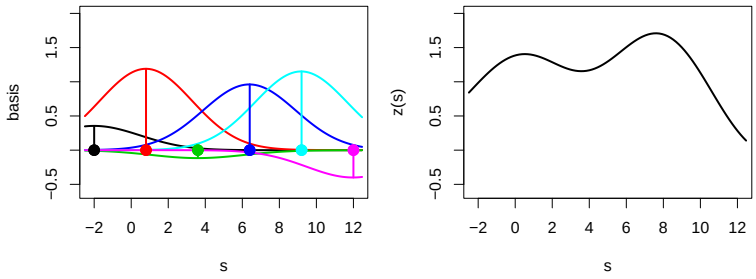


Figure 26: A spatial process realization $z(s)$ constructed from a basis representation. The basis kernels, weighted by a standard normal draw, are shown in the left plot. The resulting process $z(s)$ is the sum of these weighted kernels.

Continuing with this 1-d example, if we restrict $z(s)$ to a fine grid of n spatial locations $s_1, \ldots, s_n$ between 0 and 10, the discrete representation (9) leads to the result that

$$z \sim N(0, KK^T).$$

Here K is a $n \times m$ matrix, so that $\Sigma_z = KK^T$ is $n \times n$. Figure 27 shows the resulting covariance matrix for z as the number of points m in the latent process x varies. As m increases, the kernel is scaled down to keep the marginal variance of $z(s)$ constant. Hence if $k(s)$ denotes the kernel when $m = 6$, we use $\frac{6}{10}k(s)$ when $m = 10$ and $\frac{6}{20}k(s)$ when $m = 20$. In the top plot $\omega_1, \ldots, \omega_m$ consist of 20 equally spaced locations

between -2 and 12. In the middle plot $\omega_1, \ldots, \omega_m$ consist of 10 equally spaced locations between -2 and 12; in the bottom plot, it's $m = 6$. The resulting covariance in each case is nearly identical. This means that for any of these constructions, the induced model for the spatial process $z(s)$ is just about the same. Hence for this particular choice of kernel, m need only be as large as 6.

This convolution – or kernel – based construction of a spatial process $z(s)$ is only appropriate when a smooth representation of $z(s)$ is ok. Using a very peaked kernel ($k(s) = |s|^{-.5} \exp\{-|s|\}$, for example) will require a very dense set of support locations $\omega_1, \ldots, \omega_m$, which will make this approach impractical. Smooth kernels that have been used include the normal density and the tricube (Cleveland, 1979), which has the feature of bounded support. For a given kernel width, one would like the spacing of the ω_j's to be as spread out as possible for computational savings. However making the spacings too large can lead to unwanted "dead regions" in $z(s)$ where the process moves to 0. For the normal kernel in 1 or 2 dimensions, spacing the ω_j's no wider than the kernel sd works fine. Finally, the width of the smoothing kernel $k(s)$ controls the nature of $z(s)$ in exactly the same way the distance scaling in the covariance rule does in Section 2.

4.1 Modeling and estimation

The appeal of these convolution representations for $z(s)$ is their flexibility as a component of more general model formulations. As an example, consider recording data $y = (y(s_1), \ldots, y(s_n))^T$ at spatial locations $s_1, \ldots, s_n$. Once knot locations $\omega_j,\ j = 1, \ldots, m$ and kernel choice $k(s)$ are specified, the remaining model formulation is trivial. Assuming $y(s_i)$ is equal to the spatial process $z(s_i)$ plus an independent

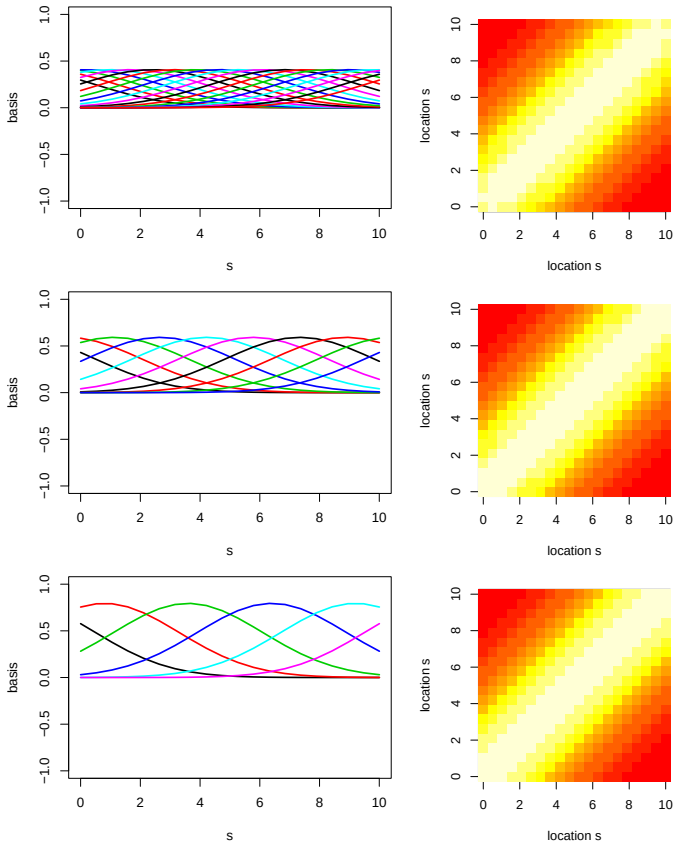


Figure 27: Basis kernels and covariance for the induced spatial process $z(s)$ as the dimension m of the latent process x varies. Here the covariance of z is nearly identical for each value of m – top: $m = 20$; middle $m = 10$; bottom: $m = 6$.

measurement error results in the likelihood

$$L(y|x, \lambda_y) \propto \lambda_y^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\lambda_y (y - Kx)^T (y - Kx)\right\}$$

where $K_{ij} = k(\omega_j - s_i)$. For priors we have

$$\begin{aligned}
\pi(x|\lambda_x) &\propto \lambda_x^{\frac{m}{2}} \exp\left\{-\tfrac{1}{2}\lambda_x x^T x\right\} \\
\pi(\lambda_x) &\propto \lambda_x^{a_x-1} \exp\{-b_x\lambda_x\} \\
\pi(\lambda_y) &\propto \lambda_y^{a_y-1} \exp\{-b_y\lambda_y\}.
\end{aligned}$$

This results in the posterior distribution

$$\begin{aligned}
\pi(x, \lambda_x, \lambda_y|y) \propto\ & \lambda_y^{a_y+\frac{n}{2}-1} \exp\left\{-\lambda_y[b_y + .5(y-Kx)^T(y-Kx)]\right\} \times \\
& \lambda_x^{a_x+\frac{m}{2}-1} \exp\left\{-\lambda_x[b_x + .5x^T x]\right\}.
\end{aligned}$$

The full conditionals then have the basic forms:

$$\begin{aligned}
\pi(x|\cdots) &\propto \exp\{-\frac{1}{2}[\lambda_y x^T K^T K x - 2\lambda_y x^T K^T y + \lambda_x x^T x]\} \\
\pi(\lambda_x|\cdots) &\propto \lambda_x^{a_x+\frac{m}{2}-1} \exp\left\{-\lambda_x[b_x + .5x^T x]\right\} \\
\pi(\lambda_y|\cdots) &\propto \lambda_y^{a_y+\frac{n}{2}-1} \exp\left\{-\lambda_y[b_y + .5(y-Kx)^T(y-Kx)]\right\}.
\end{aligned}$$

These densities are recognized as standard forms so that a Gibbs sampler can be easily implemented by cycling through the parameters and making the draws below.

$$\begin{aligned}
x|\cdots &\sim N((\lambda_y K^T K + \lambda_x I_m)^{-1}\lambda_y K^T y, (\lambda_y K^T K + \lambda_x I_m)^{-1}) \\
\lambda_x|\cdots &\sim \Gamma(a_x + \frac{m}{2}, b_x + .5x^T x) \\
\lambda_y|\cdots &\sim \Gamma(a_y + \frac{n}{2}, b_y + .5(y-Kx)^T(y-Kx))
\end{aligned}$$

As an example, we use $n = 18$ data points from spatial locations evenly spaced between 0 and 10 as shown in Figure 28. The prior model for $z(s)$ is constructed using $m = 20$ knot locations evenly spaced between -2 and 12. Finally, the kernel $k(s)$ is a normal density with a sd of 2. The resulting posterior estimate of $z(s)$ is shown in Figure 28, along

with the marginal posterior distributions for λ_y and λ_x. The following figure – Figure 29 – shows the same posterior summaries using the same data and an identical formulation, except that the kernel sd is 1 – half as wide as the one used for Figure 28. The true value for λ_y is 15 which is consistent with the formulation with the wider kernel. Note the narrow kernel used in the second formulation leads to a very wiggly estimate of $z(s)$ and a very large error precision λ_y. Hence the estimated spatial process nearly interpolates the data.

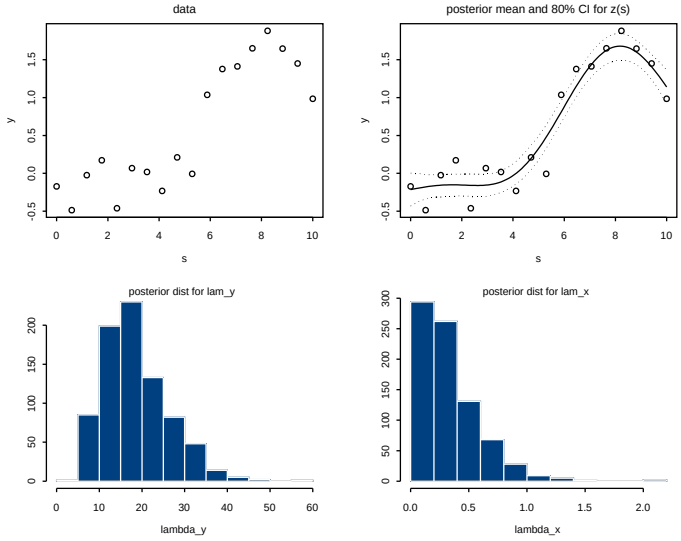


Figure 28: Data and posterior estimates. Top left: data y for estimating the underlying spatial process $z(s)$. Top right: posterior mean and pointwise 80% credible intervals for $z(s)$. Bottom: histograms of the posterior realizations for λ_y (left) and λ_x (right). Here the kernel was normal with a sd of 2.

Clearly, the kernel width can have a dramatic effect on the analysis. This width parameter could be formally incorporated into the analysis by assigning it a prior and including

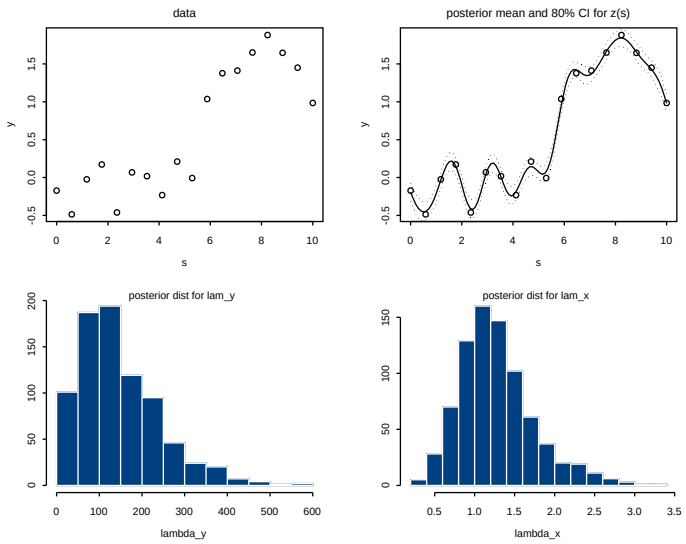


Figure 29: Data and posterior estimates. Top left: data y for estimating the underlying spatial process $z(s)$. Top right: posterior mean and pointwise 80% credible intervals for $z(s)$. Bottom: histograms of the posterior realizations for λ_y (left) and λ_x (right). Here the kernel was normal with a sd of 1 – half of that used in the previous figure.

it in the resulting MCMC. This will typically lead to a substantial amount of additional computation. Another alternative is to use a cross validation approach to settle on the kernel width. A couple of alternatives to these well established approaches for selecting the kernel width are discussed later in this section. The first is to modify the prior distribution on the latent process x; the second is to use more than one kernel in the basis construction.

4.2 Using a MRF model for x

If we take the formulation described in Section 4.1 and replace the i.i.d. prior for x with a GMRF prior

$$\pi(x|\lambda_x) \propto \lambda_x^{\frac{m}{2}} \exp\left\{-\tfrac{1}{2}\lambda_x x^T W x\right\}$$

we get the posterior

$$\begin{aligned}\pi(x, \lambda_x, \lambda_y|y) \quad \propto \quad & \lambda_y^{a_y+\frac{n}{2}-1} \exp\left\{-\lambda_y[b_y + .5(y-Kx)^T(y-Kx)]\right\} \times \\ & \lambda_x^{a_x+\frac{m}{2}-1} \exp\left\{-\lambda_x[b_x + .5x^T W x]\right\}.\end{aligned}$$

The resulting Gibbs sampler cycles through the steps

$$\begin{aligned} x|\cdots \quad &\sim \quad N((\lambda_y K^T K + \lambda_x W)^{-1}\lambda_y K^T y, (\lambda_y K^T K + \lambda_x W)^{-1}) \\ \lambda_x|\cdots \quad &\sim \quad \Gamma(a_x + \frac{m}{2}, b_x + .5x^T x) \\ \lambda_y|\cdots \quad &\sim \quad \Gamma(a_y + \frac{n}{2}, b_y + .5(y-Kx)^T(y-Kx)) \end{aligned}$$

which are no more difficult than those from the original independent x formulation.

The advantage of allowing dependence in the latent process is that the induced prior distribution for $z(s)$ is now

$$\pi(z) \propto \lambda_x^{\frac{m}{2}} \exp\{-\tfrac{1}{2}\lambda_x z^T K W K^T z\}.$$

This is an intrinsic process and λ_x controls the regularity in $z(s)$ in a way similar to kernel width. To see this, consider the posterior analyses shown in Figure 30.

Here the data are $n = 12$ observations from spatial locations evenly spaced between 0 and 10 as shown by the circle plotting symbols in Figure 30. The data were generated by adding white noise to a spatial process $z(s)$ which was constructed using $m = 20$ knot locations evenly spaced between -2 and 12, and a normal kernel $k(s)$ with a sd of 2.5.

The top plot is the result from the formulation using the iid prior for x along with a narrow kernel (sd = .6). The middle plot is the result from the formulation using the locally

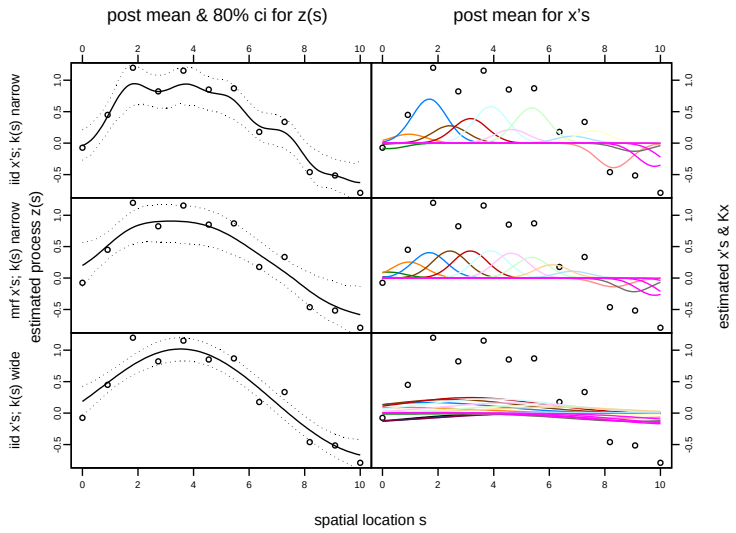


Figure 30: Posterior estimates of the underlying spatial process $z(s)$ under 3 different formulations. Top: iid prior for x with a narrow kernel (sd=.6); Middle: locally linear GMRF prior for x with a narrow kernel (sd=.6) Bottom: iid prior for x with a wide kernel (sd=2.5). The data were generated by adding noise to a spatial process generated using the wide kernel. The GMRF formulation uses the dependence in x to overcome the overly narrow specification of the kernel width. The right hand column of plots shows the posterior mean estimate of $k(s-\omega_j)x_j$ under each of the formulations.

linear GMRF prior for x along with the same narrow kernel. For the sake of comparison, the bottom plot gives the result from the formulation using the iid prior for x along with the correct, wide kernel. Clearly the formulation corresponding to the top plot overfits the data because the kernel is too narrow (by more than a factor of 4). However the GMRF formu-

lation overcomes this misspecification of the kernel width by controlling the dependence between components of x. The smoothness in the posterior reconstruction of $z(s)$ is quite similar to that of the formulation in the bottom plot which uses the correct model for the true underlying process $z(s)$ from which the data were generated.

Note that if the smoothing kernel $k(s)$ is specified to be wider than the data warrant, there is no way the MRF prior on x can overcome this misspecification. For more details on such models, see Lee *et al.*, (2005).

4.3 A multiresolution example

With these tools, it is easy to specify a spatial model which is the sum of multiple processes, each with different spatial dependence properties. To be concrete, consider the ozone measurements taken on a summer day over the Eastern U.S. shown in Figure 31. We may use a spatial process $z(s)$ to predict ozone measurements at sites where no monitors exist. Our multiresolution model decomposes the field into a coarse resolution component which varies slowly as a function of spatial distance, and a fine resolution component which changes more quickly with spatial distance

$$z(s) = z_{\text{coarse}}(s) + z_{\text{fine}}(s).$$

Separate convolution priors are used for coarse and fine spatial model components. The coarse process uses $m_c = 27$ locations $\omega_1^c, \ldots, \omega_{m_c}^c$ on a hexagonal grid as shown in Figure 31. The latent process x_c is given an independent normal prior with precision λ_c:

$$x_c = (x_{c1}, \ldots, x_{cm_c})^T \sim N(0, \frac{1}{\lambda_c} I_{m_c}).$$

The coarse kernel $k_c(s)$ is normal with a sd that is equal to the grid spacing shown in the top left plot of Figure 31.

The fine process uses $m_f = 87$ locations $\omega_1^f, \ldots, \omega_{m_f}^f$ on a hexagonal grid as shown in Figure 31. The latent process x_f is given an independent normal prior with precision λ_f:

$$x_f = (x_{f1}, \ldots, x_{fm_f})^T \sim N(0, \frac{1}{\lambda_f} I_{m_f}).$$

The fine kernel $k_f(s)$ is normal with a sd that is equal to the grid spacing shown in the bottom left plot of Figure 31.

Now the data, which consist of $n = 510$ observations, can be modeled as a sum of the two spatial processes plus white noise

$$\begin{aligned} y &= K_c x_c + K_f x_f + \epsilon \\ &= Kx + \epsilon \end{aligned}$$

where

$$K = \begin{pmatrix} K_c & K_f \end{pmatrix} \text{ and } x = \begin{pmatrix} x_c \\ x_f \end{pmatrix}.$$

If we define

$$W_x = \begin{pmatrix} \lambda_c I_{m_c} & 0 \\ 0 & \lambda_f I_{m_f} \end{pmatrix}$$

then this formulation is almost identical to previous specifications. Only now there are two precision parameters governing x. The resulting Gibbs sampler implementation is then

$$\begin{aligned} x|\cdots &\sim N((\lambda_y K^T K + W_x)^{-1} \lambda_y K^T y, (\lambda_y K^T K + W_x)^{-1}) \\ \lambda_c|\cdots &\sim \Gamma(a_x + \frac{m_c}{2}, b_x + .5 x_c^T x_c) \\ \lambda_f|\cdots &\sim \Gamma(a_x + \frac{m_f}{2}, b_x + .5 x_f^T x_f) \\ \lambda_y|\cdots &\sim \Gamma(a_y + \frac{n}{2}, b_y + .5 (y - Kx)^T (y - Kx)). \end{aligned}$$

The posterior means for the coarse resolution process $z_{\text{coarse}}(s)$ and the multiresolution process $z(s)$ are shown in Figure 31.

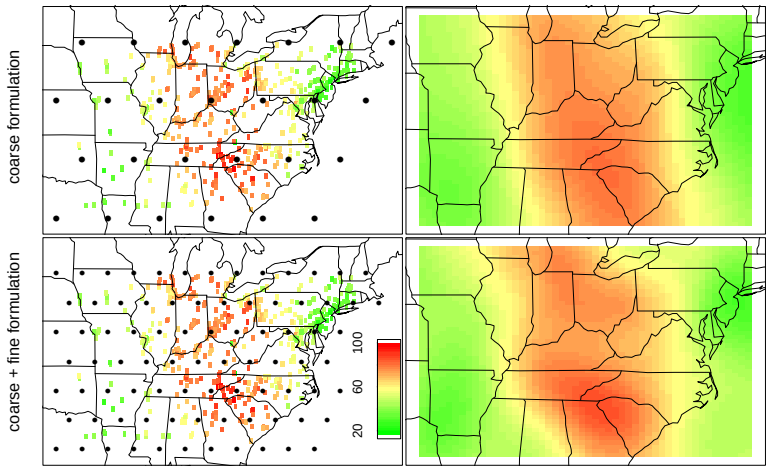


Figure 31: A multiresolution ozone model. The ozone measurements from a single summer day in the Eastern U.S. (shown in both of the left hand plots) are modeled as the sum of two process convolution models, one coarse and one fine. The knot locations for the coarse model are shown in the top left plot; the knot locations for the fine process are shown in the bottom left plot. The resulting posterior mean for $z_{\text{coarse}}(s)$ is given in the top right plot; that of the multiresolution model $z(s) = z_{\text{coarse}}(s) + z_{\text{fine}}(s)$ is shown in the bottom right plot. The kernels for the two processes are normal, each of whose sd corresponds to their respective knot spacing.

This model, which effectively uses two different kernel widths, can capture more complicated dependence structure in a spatial process $z(s)$. It also can result in a model with a single kernel width if the data support this by moving the posterior distribution for one of λ_c or λ_f to a very large value. This approach has links to wavelets (see Vidakovic (1999)) and could be modified to have more resolutions if the application calls for it.

4.4 A binary spatial model

Here we show how these convolution-based models are quite useful for modeling non-Gaussian spatial fields. Here we focus on a binary spatial model, although these ideas could be extended to other non-Gaussian fields as well. This binary model borrows from the ideas of De Olivaira (2000).

Consider an example shown in Figure 32 where there exists a binary field $z^*(s)$ over $\mathcal{S} = [0, 20] \times [0, 20]$. We obtain

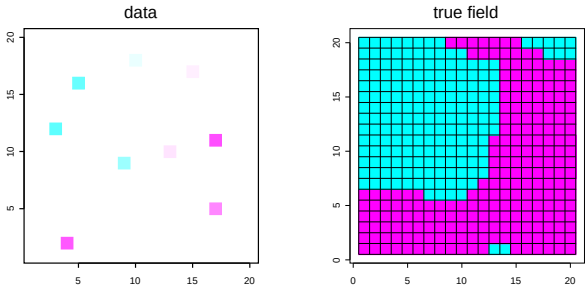


Figure 32: Observations from a binary field. The left plot shows $n = 9$ noisy measurements from the true binary field $z^*(s)$ shown on the right. The blue denotes $z^*(s) = 1$; red denotes $z^*(s) = 0$. The datapoints $y(s_i)$ are created by adding a $N(0, 1)$ error to the true field $z^*(s_i)$.

$n = 9$ noisy measurements $y = (y_1, \ldots, y_n)^T$ of this field at spatial locations $s_1, \ldots, s_n$ shown in the figure. The sampling model is

$$y_i = z^*(s_i) + \epsilon_i, \; i = 1, \ldots, n, \; \text{where } \epsilon \sim N(0, I_n).$$

The binary field $z^*(s)$ is modeled as a Gaussian field $z(s)$, constructed via process convolution, and then assigning $z^*(s) = I[z(s) > 0]$ as shown in Figure 33. The knot locations are a 5×5 lattice shown in Figure 33 so that $m = 25$; the smoothing kernel is Gaussian and is also shown in the

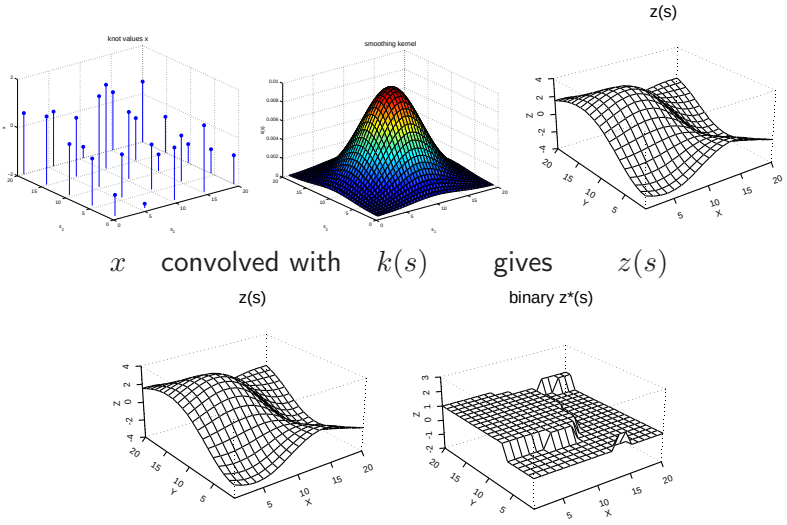


Figure 33: Constructing a binary field $z^*(s)$ by clipping a Gaussian field $z(s)$. The Gaussian process $z(s)$ is constructed by smoothing out a lattice of iid normal realizations x. This resulting field is then "clipped" so that $z^*(s) = I[z(s) > 0]$.

figure. Take z^* to denote the value of the binary process at the data locations $z^* = (z^*(s_1), \ldots, z^*(s_n))^T$ and recall that z^* is a deterministic function of the m-vector x.

The resulting posterior distribution for this formulation is simply the product of the likelihood

$$L(y|z^*) \propto \exp\{-\tfrac{1}{2}(y - z^*)^T(y - z^*)\}$$

and the independent normal prior for x

$$\pi(x) \propto \exp\{-\tfrac{1}{2}x^T x\}$$

giving

$$\pi(x|y) \propto \exp\{-\tfrac{1}{2}(y - z^*)^T(y - z^*) - \tfrac{1}{2}x^T x\}.$$

The full conditional distributions resulting from this formulation are

$$\pi(x_j|x_{-j},y) \propto \exp\{-\tfrac{1}{2}(y-z^*)^T(y-z^*)-\tfrac{1}{2}x_j^2\},\ j=1,\ldots,m.$$

Because z^* is a nonlinear function of x, this density does not have a recognizable form. Hence a Gibbs sampler is not an easy option here. A simple alternative is to use Metropolis updating (Metropolis *et al.*, 1953) to generate a MCMC sample from this posterior.

A Metropolis implementation for sampling from $\pi(x|y)$ can be carried out by first initializing x at 0 and then cycling thru full conditionals updating each x_j in turn, according to the Metropolis rules:

- generate proposal $x_j' \sim U[x_j - r, x_j + r]$;
- compute acceptance probability

$$\alpha = \min\left\{1, \frac{\pi(x_j'|x_{-j},y)}{\pi(x_j|x_{-j},y)}\right\};$$

- update x_j to new value:

$$x_j^{\text{new}} = \begin{cases} x_j' & \text{with probability } \alpha \\ x_j & \text{with probability } 1-\alpha; \end{cases}$$

Here the sampler ran for $T = 1000$ scans, giving realizations $x^1, \ldots, x^T$ from the posterior. The output from the first 100 scans were discarded for burn in. Note, we tuned the proposal width r so that the proposal x_j' is accepted about half the time.

A subset of posterior realizations for $z^*(s)$ are shown in Figure 34. There is clearly a wide range of configurations for $z^*(s)$ that are consistent with the data. The resulting posterior mean estimate for $z^*(s)$ is shown in Figure 35. The process convolution formulation for $z^*(s)$ works very well with the Metropolis MCMC implementation. One reason for this

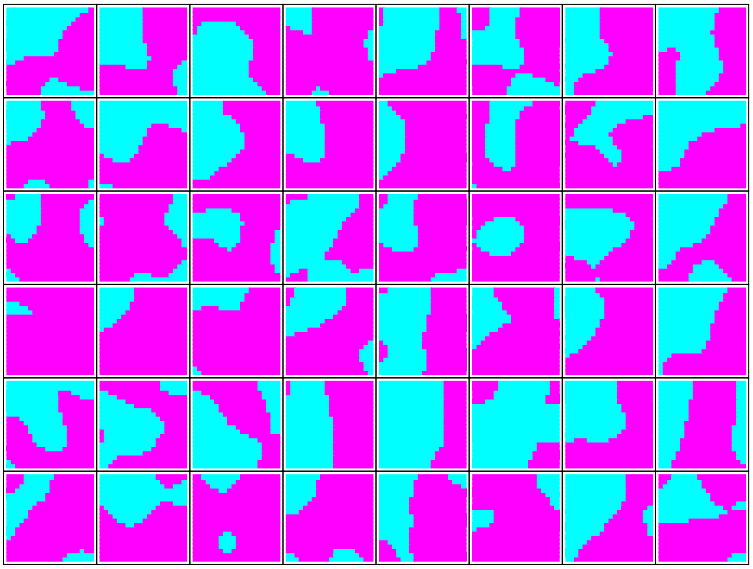


Figure 34: Posterior realizations of the binary process $z^*(s)$.

is that the individual x_j's control just a local piece of the entire process $z^*(s)$. Hence updates to a x_j are effectively updates to a local region of $z^*(s)$.

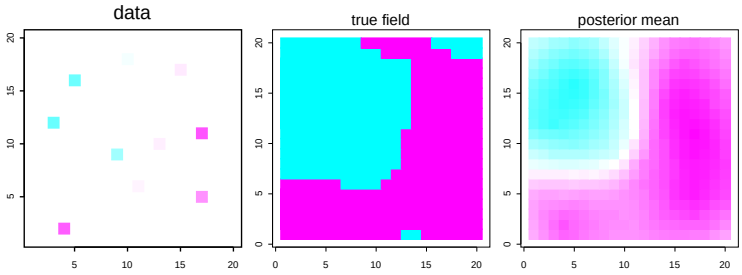


Figure 35: Data (left), true binary field $z^*(s)$ (middle), and posterior mean estimate of $z^*(s)$ (right). The posterior mean estimate is simply the average of the MCMC draws of $z^*(s)$.

4.4.1 Archeology application

This same model can be applied to the Archeology application of Besag *et al.*, (1991). In this application, as well as many other archeological investigations, enhanced soil phosphate content, due to decomposition of organic mater, is often found at known locations of archeological activity. In such areas, measurements of phosphate concentration can help aid in determining specific sites of archeological activity.

The application centers on a square plot divided into a 16 $\times$ 16 grid of sites, each measuring 10 meters by 10 meters; The layout and data are shown in Figure 36. As in the earlier analysis, we assume the existence of an underlying process, $z^*(s)$, with $z^*(s_i) = 1$ or 0, according to whether there is or is not previous activity at site s_i. The goal here is not only to produce a classification, but also to explore the posterior distribution of $z^*(s)|y$ and to determine posterior probability of previous activity at each site.

The sampling model is adapted from the original application and has the form:

$$y(s_i)|z^*(s_i) \sim N(\mu(z^*(s_i)), 1) \text{ where } \mu(0) = 1,\ \mu(1) = 2.$$

Recall $z(s) = \sum_{j=1}^{m} x_j k(s - \omega_j)$ and $z^*(s) = I[z(s) > 0]$. A $m = 8 \times 8$ lattice of knot locations $\omega_1, \ldots, \omega_m$ is used as shown in Figure 36. The sd of $k(s)$ is equal to the minimum knot spacings. A brief posterior summary for $z^*(s)$ is given in Figure 36 which shows posterior realizations of $z^*(s)$ along with its posterior mean.

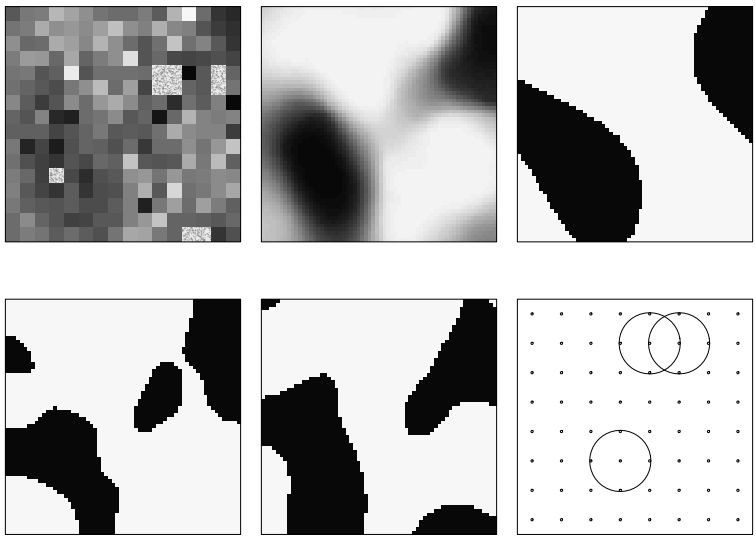


Figure 36: Data, posterior mean, posterior realizations, and knot locations for the archeology application of Besag *et al.*, (1991). Here, the binary process is a "clipped" Gaussian field constructed from a process convolution model. The bottom right frame shows the knot locations and a circle that corresponds to 1 sd of the bivariate smoothing kernel $k(s)$.

Historical note on Metropolis

Nick Metropolis (pictured in Figure 37) was involved with the original Manhattan Project at Los Alamos during World War II. He, along with John von Neumann, were computing pioneers at Los Alamos. He is the inventor of the Monte Carlo Method as well as MCMC, which was first carried out in 1953 on the MANIAC I computer (also pictured in Figure 37). This machine was the first computer that could carry out instructions based on previous computations. The MANIAC I also ran the first computer game – chess.

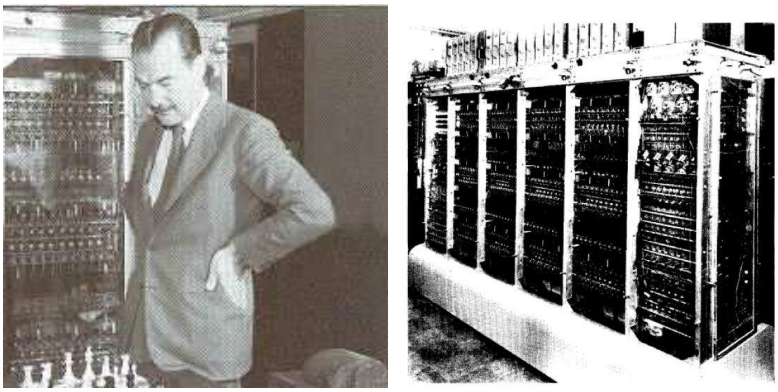


Figure 37: Left: Nick Metropolis, a computing pioneer and inventor of the Monte Carlo method. Right: the MANIAC I computer.

5 Convolution-based space-time models

The convolution-based models can easily be extended in two basic ways to model space-time processes. The first simply defines the latent process $x(s,t)$ to reside on knot values taken over a lattice in space and time, and then defines the kernel $k(s,t)$ to smooth over this augmented space. Mathematically, this approach is the same as that of the previous section, only an additional dimension is added. The second approach allows the latent spatial process to evolve over time, so that at any given time, it can be convolved with the spatial kernel $k(s)$ to give a spatial field $z(s,t)$ at time t. Both of these examples will be demonstrated with an application.

5.1 A space-time convolution model

The most straightforward extension of the spatial convolution specification to space-time models extends the spatial domain to a space-time domain $\mathcal{S}\times\mathcal{T}$. The latent knot process $x(s,t)$ is defined over m points $\{(\omega_1,\tau_1),\ldots,(\omega_m,\tau_m)\}$ within $\mathcal{S}\times\mathcal{T}$. In addition, the smoothing kernel $k(s,t)$ is also defined over the space-time domain. The space-time process $z(s,t)$ is now defined as

$$\begin{aligned} z(s,t) &= \int_{\mathcal{S}\times\mathcal{T}} k((\omega,\tau)-(s,t))dx(\omega,\tau) \\ &= \sum_{j=1}^{m} k((\omega_j,\tau_j)-(s,t))x(\omega_j,\tau_j) \\ &= \sum_{j=1}^{m} k_{st}(\omega_j,\tau_j)x_j \end{aligned}$$

where $x_j = x(\omega_j,\tau_j),\ j=1,\ldots,m$. This constructive model is shown in Figure 38.

We now use this model to build an interpolative space time model $z(s,t)$ for ocean temperatures on a manifold of

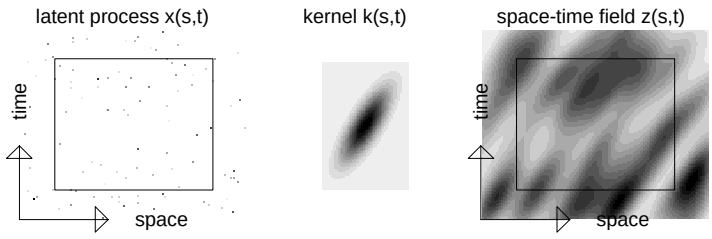


Figure 38: A convolution based space-time model. Here a space time-model $z(s,t)$ (right) is constructed by defining a latent process $x(s,t)$ over space and time (left) and convolving it with the kernel $k(s,t)$ (middle).

constant potential density. This application is taken from Higdon (1998). The data consist of $n = 3987$ temperature measurements taken between 1908 and 1988 shown in Figure 39. We take $y = (y_1, \ldots, y_n)^T$ to denote the recorded temperatures at irregularly sampled locations $(s_1, t_t), \ldots, (s_n, t_n)$. We center the data so that $\bar{y} = 0$. We assume a simple model for which the data are equal to a smooth process $z(s,t)$ plus white noise.

The smooth space-time process $z(s,t)$ is constructed by smoothing a latent space-time process x with a spatially varying kernel $k_s(s,t)$. The latent process resides on a regular grid over space and time. The spatial locations are shown in the right hand plot of Figure 39. The kernel $k_{st}(\omega, \tau)$ at any given spatial location has the product form $k_s(\omega)k_t(\tau)$ where $k_s(\cdot)$ varies with with spatial location and and $k_t(\cdot)$ is a fixed, 1-d normal kernel with a sd of 7 years. This product form for $k_{st}(\omega, \tau)$ implies a separable covariance rule that is the product of a purely temporal covariance rule and a purely spatial covariance rule. The spatially varying kernel $k_s(\cdot)$ is bivariate normal. The 1 sd ellipse is shown for 8 spatial locations in

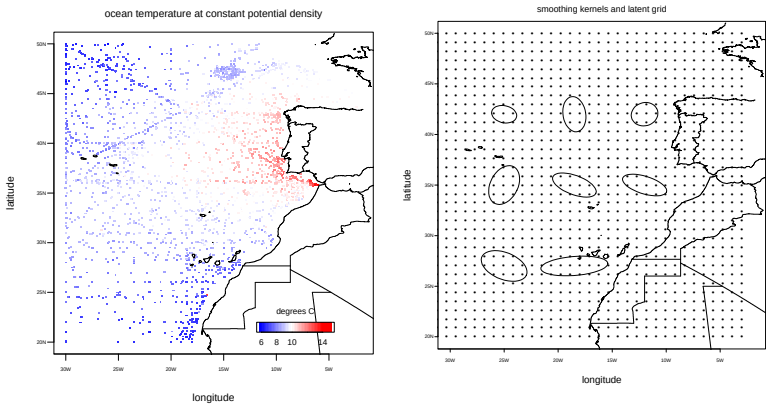


Figure 39: Left: Measured temperatures between 1908 and 1988 along a manifold of constant potential density, which corresponds to depths well below 1000m. The data consist of 3987 measurements. Right: Spatial locations of the latent space-time process $x(s,t)$. The time spacings are every seven years. Also shown are ellipses that correspond to one sd of the spatially varying gaussian smoothing kernels $k_s(\cdot)$.

Figure 39. Hence, the space-time process $z(s,t)$ is given by

$$z(s,t) = \sum_{j=1}^{m} k_{st}(\omega_j, \tau_j) x_j.$$

The details of estimating and specifying the kernel is given in Higdon (1998).

The resulting formulation is summarized below.

Likelihood:

$$L(y|x,\lambda_y) \propto \lambda_y^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\lambda_y (y-Kx)^T (y-Kx)\right\}$$

where $K_{ij} = k_{s_i t_i}(\omega_j, \tau_j)$.

Priors:

$$\begin{aligned}
\pi(x|\lambda_x) &\propto \lambda_x^{\frac{m}{2}} \exp\left\{-\frac{1}{2}\lambda_x x^T x\right\} \\
\pi(\lambda_x) &\propto \lambda_x^{a_x - 1} \exp\{-b_x \lambda_x\} \\
\pi(\lambda_y) &\propto \lambda_y^{a_y - 1} \exp\{-b_y \lambda_y\}
\end{aligned}$$

Posterior:

$$\begin{aligned}
\pi(x, \lambda_x, \lambda_y | y) \propto\ & \lambda_y^{a_y + \frac{n}{2} - 1} \exp\left\{-\lambda_y [b_y + .5(y-Kx)^T (y-Kx)]\right\} \times \\
& \lambda_x^{a_x + \frac{m}{2} - 1} \exp\left\{-\lambda_x [b_x + .5 x^T x]\right\}
\end{aligned}$$

This posterior is exactly in the same form as the analogous posterior from the purely spatial convolution formulations.

The Gibbs sampler implementation is then

$$
\begin{aligned}
x|\cdots &\sim N((\lambda_y K^T K + \lambda_x I_m)^{-1}\lambda_y K^T y, (\lambda_y K^T K + \lambda_x I_m)^{-1}) \\
x_j|\cdots &\sim N\left(\frac{\lambda_y r_j^T k_j}{\lambda_y k_j^T k_j + \lambda_x}, \frac{1}{\lambda_y k_j^T k_j + \lambda_x}\right) \\
\lambda_x|\cdots &\sim \Gamma(a_x + \frac{m}{2}, b_x + .5x^T x) \\
\lambda_y|\cdots &\sim \Gamma(a_y + \frac{n}{2}, b_y + .5(y - Kx)^T(y - Kx))
\end{aligned}
$$

where k_j is jth column of K and $r_j = y - \sum_{j' \neq j} k_{j'} x_{j'}$. Note that given the large number of knots used here, the latent process will most likely need to be updated using the single site updating corresponding to the second line above. The resulting posterior mean for $z(s,t)$ is shown for selected times in Figure 40. For an alternative formulation of a space-time model for this data see Stroud *et al.*(1999).

5.2 A spatial convolution of a temporally evolving latent process

Rather than rely on the kernel $k_{st}(\omega, \tau)$ to induce the temporal dependence in $z(s,t)$, an appealing alternative is to put the temporal dependence in the latent process $x(s,t)$. The space-time field $z(s,t)$ can then be obtained by a purely spatial convolution at time t. This is depicted in Figure 41. The advantage here is that the convolution is purely spatial, which can lead to substantial computational savings. Such an approach is typically only practical when observations come in at regular times.

Here the space-time process $z(s,t)$ is defined on the domain $\mathcal{S} \times \mathcal{T}$ where $\mathcal{T} = 1, 2, \ldots, m_t$. The discrete latent process $x(\omega, t)$ is defined at space-time knot locations $\{\omega_1, \ldots, \omega_{m_s}\} \times \{1, \ldots, m_t\}$, where $m_t \cdot m_s = m$. Finally, a spatial smoothing

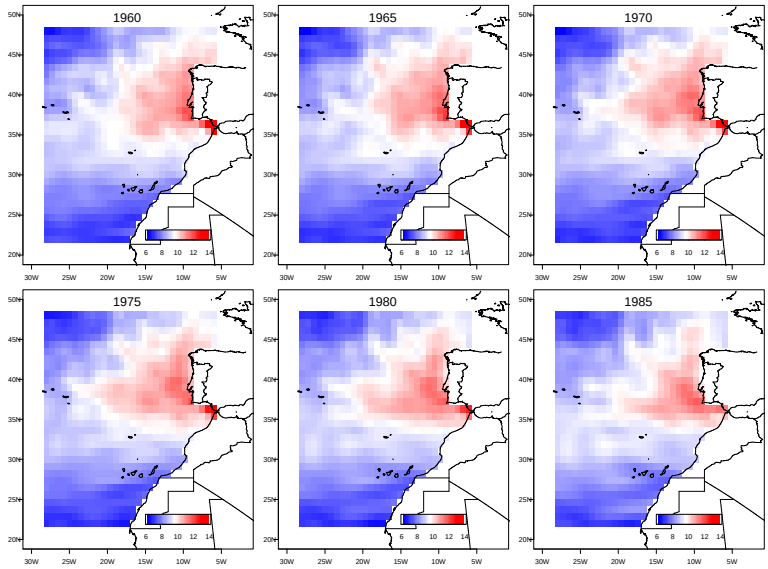


Figure 40: Posterior mean estimate of the temperature surface $z(s,t)$ at six different times.

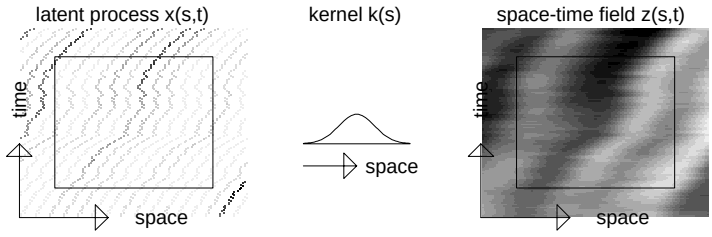


Figure 41: A convolution based space-time model. Here a space time-model $z(s,t)$ (right) is constructed by defining a latent process $x(s,t)$ that is evolving over time (left) and convolving it with a purely spatial kernel $k(s)$ (middle).

kernel $k(s)$ is defined to smooth out the latent process separately for each time. More generally, a family of smoothing kernels $k_{st}(\omega)$ may be defined that vary with each space-time point (s,t). The space-time process is then constructed as defined below.

$$z(s,t) \quad = \quad \sum_{j=1}^{m_s} k(\omega_j - s)x(\omega_j, t)$$

If a kernel that varies with space and time is used, the above construction generalizes to

$$z(s,t) \quad = \quad \sum_{j=1}^{m_s} k_{st}(\omega_j)x_{jt}$$

where x_{jt} is shorthand for $x(\omega_j, t)$. In the upcoming ozone example, a single, circular, bivariate Gaussian kernel is used for $k(s)$.

The spatial dependence is incorporated into the prior distribution for the latent process $x(\omega, t)$. Arguably, the simplest extension to the independence model is to define a locally linear GMRF prior for each $x_j = (x_{j1}, \ldots, x_{jm_t})^T$. Thus

$$\pi(x_j|\lambda_x) \propto \lambda^{\frac{m_t}{2}} \exp\left\{-\frac{1}{2}\lambda_x x_j^T W x_j\right\}$$

where

$$W_{ij} = \begin{cases} -1 & \text{if } |i-j| = 1 \\ 1 & \text{if } i = j = 1 \text{ or } i = j = m_t \\ 2 & \text{if } 1 < i = j < m_t \\ 0 & \text{otherwise.} \end{cases}$$

So for the entire latent process

$$x = (x_{11}, x_{21} \ldots, x_{m_s 1}, x_{12}, \ldots, x_{m_s 2}, \ldots, x_{1m_t}, \ldots, x_{m_s m_t})^T$$

the prior density becomes

$$\pi(x|\lambda_x) \propto \lambda^{\frac{m_t m_s}{2}} \exp\left\{-\frac{1}{2}\lambda_x x^T (W \otimes I_{m_s}) x\right\}.$$

This formulation is applied to a sequence of 60 days worth of ozone measurements – 5 of which are shown in Figure 42. Here the data $y = (y_1, \ldots, y_n)^T$ consist of $n_s = 510$ readings

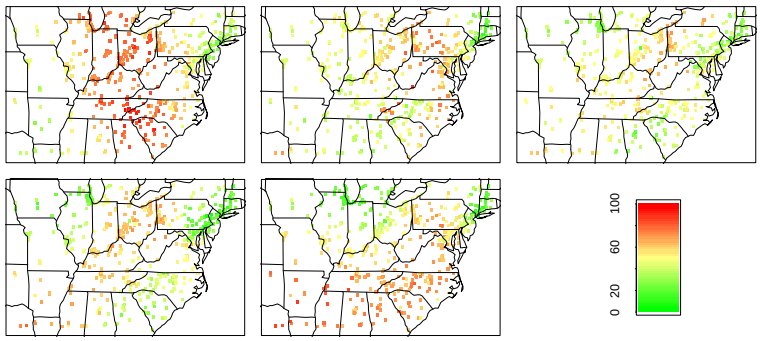


Figure 42: Ozone concentrations taken for 5 consecutive summer days over the eastern US. This is a subset of the 60 days of measurements used in the example.

taken daily at spatial locations $s_1, \ldots, s_{n_s}$, for $m_t = 60$ consecutive days so that $n = 510 \cdot 60$. For a single day t, the likelihood for the n_s-vector of observations $y_t = (y_{1t}, \ldots, y_{nt})^T$

is

$$L(y_t|x_t,\lambda_y) \propto \lambda_y^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\lambda_y(y_t - K^t x_t)^T(y - K^t x_t)\right\}$$

where $K_{ij}^t = k(\omega_j - s_i)$. For the entire set of n observations $y = (y_1^T, \ldots, y_{n_t}^T)^T$, the likelihood is

$$L(y|x,\lambda_y) \propto \lambda_y^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\lambda_y(y - Kx)^T(y - Kx)\right\}$$

where $K = \text{diag}(K^1, \ldots, K^{n_t})$.

The prior for x is determined by the $m_s = 27$ knot locations and spatial smoothing kernel used in the coarse model formulation shown in the top left plot of Figure 31. For the duration of $m_t = 60$ days, the total number of knots is $m = m_s \cdot m_t = 27 \cdot 60$. So the full prior specification can be written

$$\begin{aligned}
\pi(x|\lambda_x) &\propto \lambda_x^{\frac{m}{2}} \exp\left\{-\frac{1}{2}\lambda_x x^T W x\right\} \\
\pi(\lambda_x) &\propto \lambda_x^{a_x - 1} \exp\{-b_x\lambda_x\} \\
\pi(\lambda_y) &\propto \lambda_y^{a_y - 1} \exp\{-b_y\lambda_y\}.
\end{aligned}$$

where $a_x = a_y = 1$ and $b_x = b_y = .001$. This results in the posterior density

$$\begin{aligned}
\pi(x,\lambda_x,\lambda_y|y) \propto\ & \lambda_y^{a_y + \frac{n}{2} - 1} \exp\left\{-\lambda_y[b_y + .5(y - Kx)^T(y - Kx)]\right\} \times \\
& \lambda_x^{a_x + \frac{m}{2}} \exp\left\{-\lambda_x[b_x + .5x^T W x]\right\}.
\end{aligned}$$

The full conditionals densities are

$$\begin{aligned}
\pi(x|\cdots) &\propto \exp\{-\frac{1}{2}[\lambda_y x^T K^T K x - 2\lambda_y x^T K^T y + \lambda_x x^T W x]\} \\
\pi(\lambda_x|\cdots) &\propto \lambda_x^{a_x + \frac{m}{2} - 1} \exp\left\{-\lambda_x[b_x + .5x^T W x]\right\} \\
\pi(\lambda_y|\cdots) &\propto \lambda_y^{a_y + \frac{n}{2} - 1} \exp\left\{-\lambda_y[b_y + .5(y - Kx)^T(y - Kx)]\right\}.
\end{aligned}$$

All of which have standard forms so that the Gibbs sampler implementation cycles through the draws

$$\begin{aligned} x|\cdots &\sim N((\lambda_y K^T K + \lambda_x W)^{-1}\lambda_y K^T y, (\lambda_y K^T K + \lambda_x W)^{-1}) \\ &\quad \text{or} \\ x_{jt}|\cdots &\sim N\left(\frac{\lambda_y r_{tj}^T k_{tj} + n_j \bar{x}_{\partial j}}{\lambda_y k_{tj}^T k_{tj} + n_j \lambda_x}, \frac{1}{\lambda_y k_{tj}^T k_{tj} + n_j \lambda_x}\right) \\ \lambda_x|\cdots &\sim \Gamma(a_x + \frac{m}{2}, b_x + .5x^T x) \\ \lambda_y|\cdots &\sim \Gamma(a_y + \frac{n}{2}, b_y + .5(y - Kx)^T (y - Kx)) \end{aligned}$$

where k_{tj} is jth column of K^t, $r_{tj} = y_t - \sum_{j' \neq j} k_{tj'} x_{tj'}$, $n_j =$ number of neighbors of x_{jt}, and $\bar{x}_{\partial jt}$ is the mean of neighbors of x_{jt}. The full conditionals are for both the multivariate and single site updates of x. For large problems the multivariate update of x is not feasible, while the univariate update of each x_{jt} will be simple regardless of how large m is. The posterior mean estimate for the space-time field $z(s,t),\ t = 1, \ldots, 5$, is shown in Figure 43. Note that this particular formulation is amenable to the dynamic linear model framework of West and Harrison (1997) which can also ease the computational burden of exploring the posterior. Here the observation and evolution equations are

$$y_t = K^t x_t + \epsilon_t,\ x_t = x_{t-1} + \varepsilon_t$$

where ϵ_t and ε_t are spatially and temporally uncorrelated. Updating x given the precisions can be carried out efficiently via forward filtering backward simulation.

6 Combining simulations and experimental observations

It is common that computer codes based on mathematical descriptions of the physical process over space and time exist to

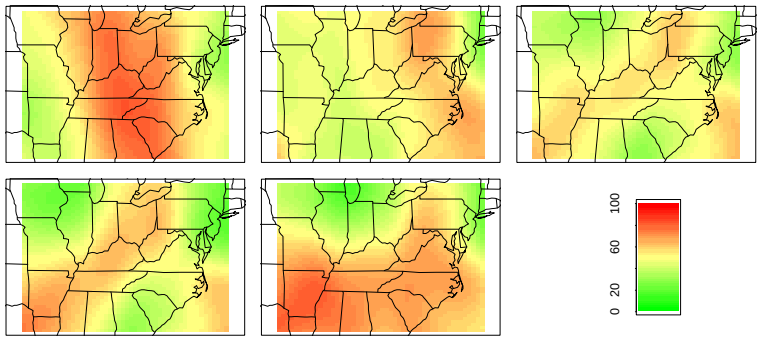


Figure 43: Posterior mean field for ozone concentration for the 5 days shown in Figure 42.

simulate the system of interest. Such simulations can account for rich space-time dependence without the need to cook up a stochastic description of the covariance in such a process. In this section we'll consider a very simple space-time system from Lorenz (1996). Given data generated from this model, we'll estimate the field at unobserved space-time locations using three different approaches: a Gaussian process model estimated from simulation output; a Bayesian inverse modeling approach; and a MRF model motivated by the partial differential equation (PDE) system itself.

Estimated fields derived from simulations of actual physical processes typically contain uncertainties due to:

- observation noise;
- uncertain initial/boundary conditions;
- uncertain coefficients/parameters in the mathematical description; and
- inadequate mathematical representation of the physical system being modeled.

Hence, given observations of this process, there will necessarily be uncertainty regarding what this process is at unobserved space-time locations.

6.1 The L96 model

We consider a deterministic, nonlinear system from Lorenz (1996) which is a continuous time model defined over discrete, periodic spatial locations $s \in \{1, 2, \ldots, 40\}$

$$\frac{\partial}{\partial t} z(s,t) = [z(s+1,t) - z(s-2,t)]z(s-1,t) - z(s,t) + F.$$

Given $z(s, t = 0)$, the system can be propagated forward via Euler integration:

$$z(s,t+dt) = z(s,t) + dt\{[z(s+1,t) - z(s-2,t)]z(s-1,t) - z(s,t) + F\}$$

–of course, more sophisticated integration schemes will give more efficient simulations. To ensure stability of the simulation, we use time increments of $dt = .001$. The forcing term F is set to 8. Figure 44 shows realizations from this model given an initial field $z(s, t = 0)$ at $n_t = 51$ regularly spaced time increments of $\Delta t = .02$, starting at $t = 0$ and ending at $t = 1$.

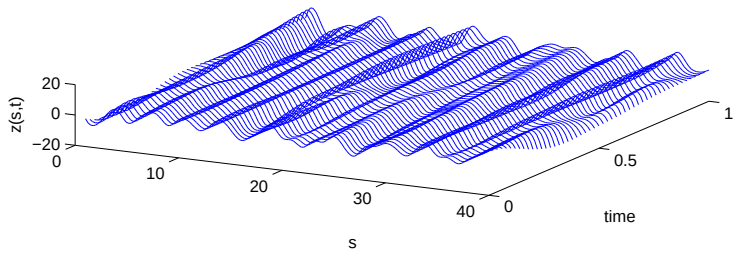


Figure 44: Output from the L96 model given an initial field $z(s, 0)$ at regularly spaced time increments $t = 0, .02, \ldots, 1$.

We generate a "true" realization $z(s,t)$ by starting with an initial field $z(s,0)$ at time $t=0$ and propagating it until $t=1$ is reached. We restrict ourselves to the space-time set $\mathcal{S}\times\mathcal{T}=\{1,2,\ldots,40\}\times\{0,.02,\ldots,1\}$ so that the true value $z(s,t)$ can be thought of as a $n_s\times n_t=40\times 51$ image shown in Figure 45. From this ground truth, we obtain $n=79$ noisy measurements

$$y_{d_i}=z(s_{d_i},t_{d_i})+\epsilon_{d_i},\ i=1,\ldots,n$$

where the measurement error $\epsilon\sim N(0,\frac{1}{4}I_n)$. The measurement locations (s_{d_i},t_{d_i}) are also shown in Figure 45. We'll

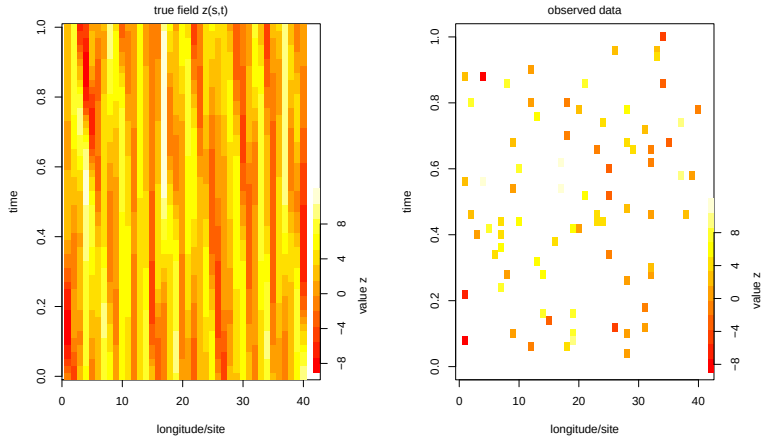


Figure 45: The true field $z(s,t)$ generated from the L96 model (left) and the noisy data (right) from which we will estimate $z(s,t)$. The observation errors are independent $N(0,\frac{1}{4})$ random deviates.

assume that we have a good estimate of the error precision. Hence we'll either treat λ_y as known ($\lambda_y=4$), or we'll use an informative $\Gamma(4,1)$ prior for λ_y which has mean 4.

If we take z to be the n_sn_t-vector containing the entire field, we can define T to be a $n\times n_sn_t$ incidence matrix that

assigns observation sites d_i to lattice locations j. Thus the likelihood can be written

$$L(y|z) \propto \lambda_y^{\frac{n}{2}} \exp\{-\tfrac{1}{2}\lambda_y (y - Tz)^T (y - Tz)\}.$$

6.2 A Gaussian process model

We can use simulations from the L96 model to obtain estimates of the mean and covariance of $z(s,t)$ over the $n_s \times n_t$ lattice. This is easily accomplished by taking a long simulation and "cutting" it into pieces of size $n_s \times n_t$. Each of these pieces can be thought of as draws $z_1, z_2, \ldots, z_M$ from a prior model for $z(s,t)$.

We can then use these realizations to build a GP prior for $z(s,t)$ over a $n_s \times n_t$ lattice over space and time

$$z \sim N(\mu_z, \Sigma_z).$$

The mean is set to the sample mean of the draws

$$\mu_z = \frac{1}{M} \sum_{j=1}^{M} z_j.$$

Here μ_z is sufficiently close to 0 that we set it to 0. Then the $n_s n_t \times n_s n_t$ covariance matrix is set to

$$\Sigma_z = \frac{1}{M} \sum_{j=1}^{M} (z_j - \mu_z)(z_j - \mu_z)^T.$$

Hence we have exactly the setup we had before in the soft conditioning case of Section 2.

$$\begin{aligned} L(y|z) &\propto \lambda_y^{\frac{n}{2}} \exp\{-\frac{1}{2}\lambda_y (y - Tz)^T (y - Tz)\} \\ \pi(z) &\propto |\Sigma_z|^{-\frac{1}{2}} \exp\{-\frac{1}{2} z^T \Sigma_z^{-1} z\} \end{aligned}$$

Therefore we know the posterior distribution for z has the form

$$z|y \sim N(\Lambda^{-1}\lambda_y T^T y, \Lambda^{-1})$$

where $\Lambda = (\lambda_y T^T T + \Sigma_z^{-1})$.

The posterior mean field is shown in Figure 46. Also shown is the residual field (true value of z - posterior mean). The root mean square (RMS) error of the residuals is 3.03. Figure 47 shows the fitted values against the true field. Also shown in this Figure are the pointwise 80% credible intervals for the initial field $z_0 = (z(1,0), z(2,0), \ldots, z(n_s,0))^T$. This initial field is of interest because it is the uncertain initial condition that determines the rest of the space-time field. The next approach focuses on estimating this initial condition z_0.

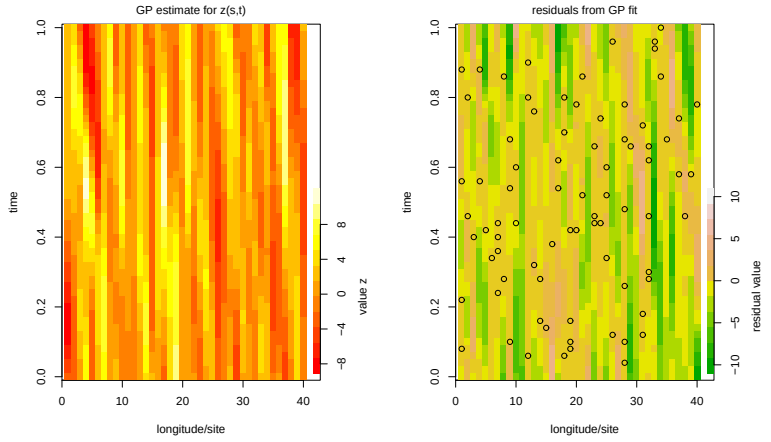


Figure 46: Left: Posterior mean for $z(s,t)$ estimated under the Gaussian process formulation. Right: The residual field obtained by subtracting the posterior mean estimate from the true field shown in Figure 45. The circles denote locations where measurements were obtained.

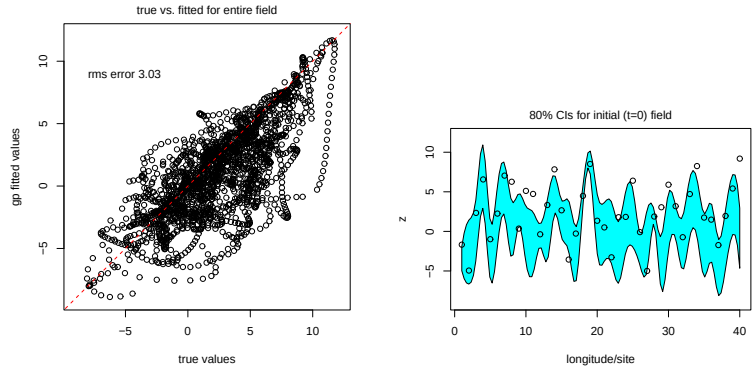


Figure 47: Left: True $z(s,t)$ plotted against the posterior mean from the Gaussian process formulation. Right: Pointwise 80% credible intervals for the initial condition $z_0 = (z(1,0), z(2,0), \ldots, z(n_s,0))^T$. The circles denote the true values for z_0.

6.3 Bayesian inverse formulation

The previous GP formulation only uses the simulator to estimate the mean and covariance of the GP model. Presumably, the simulation model can be much more informative about what fields are plausible given the observed data. An approach that makes much stronger use of the simulator is the Bayesian inverse formulation. This comes at a price – a huge number of simulation runs will be required.

The initial state for the space-time field $z(s,t)$ is given by the n_s-vector $z_0 = (z(1,0), \ldots, z(40,0))^T = z(s, t = 0)$. Given z_0, the remaining field z is determined by the L96 model

$$z = \eta(z_0)$$

where the simulator output z is defined over the $n_s \times n_t = 40 \times 51$ lattice. Hence the likelihood – which requires an L96

simulation to evaluate – is

$$L(y|z_0,\lambda_y) \propto \lambda_y^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\lambda_y (y - T\eta(z_0))^T (y - T\eta(z_0))\right\}.$$

A locally linear GMRF prior is specified for z_0:

$$\pi(z_0|\lambda_0) \propto \lambda_0^{\frac{n_s}{2}} \exp\left\{-\tfrac{1}{2}\lambda_0 \sum_{i=1}^{n_s} (z_{0i} - z_{0i+1})^2\right\}.$$

Note the sum goes to n_s here since the spatial sites are assumed to be periodic. In the sum n_s+1 taken to be 1, making sites 1 and n_s neighbors. The prior specification is completed by specifying priors for the precision parameters:

$$\begin{aligned}
\pi(\lambda_y) &\propto \lambda_y^{a_y-1} \exp\{-b_y\lambda_y\}, \quad a_y = 4; \quad b_y = 1, \\
\pi(\lambda_0) &\propto \lambda_0^{a_0-1} \exp\{-b_0\lambda_0\}, \quad a_0 = 1; \quad b_0 = .001.
\end{aligned}$$

This results in the posterior distribution:

$$\begin{aligned}
\pi(z_0,\lambda_y,\lambda_0|y) &\propto L(y|\eta(z_0),\lambda_y) \times \pi(z_0|\lambda_0) \times \pi(\lambda_y) \times \pi(\lambda_0) \\
&\propto \lambda_y^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\lambda_y (y - T\eta(z_0))^T (y - T\eta(z_0))\right\} \times \\
&\quad \lambda_0^{\frac{n_s}{2}} \exp\left\{-\frac{1}{2}\lambda_0 \sum_{i=1}^{n_s-1} (z_{0i} - z_{0i+1})^2\right\} \times \\
&\quad \lambda_y^{a_y-1} \exp\{-b_y\lambda_y\} \times \lambda_0^{a_0-1} \exp\{-b_0\lambda_0\}.
\end{aligned}$$

The resulting full conditionals are then

$$\begin{aligned}
\pi(z_{0i}|\cdots) &\propto \exp\left\{-\frac{1}{2}\lambda_y (y - T\eta(z_0))^T (y - T\eta(z_0))\right\} \times \\
&\quad \exp\left\{-\frac{1}{2}\lambda_0 \sum_{j\in\partial i} (z_{0i} - z_{0j})^2\right\} \\
\lambda_y|\cdots &\sim \Gamma(a_y + n/2, b_y + .5(y - T\eta(z_0))^T (y - T\eta(z_0))) \\
\lambda_0|\cdots &\sim \Gamma\left(a_0 + n_s/2, b_0 + .5\sum_{i=1}^{n_s} (z_{0i} - z_{0i+1})^2\right).
\end{aligned}$$

Note that evaluating $\pi(z_{0i}|\cdots)$ involves running the L96 simulator. Also this full conditional will require a Metropolis update in the MCMC, while the precisions can be updated via Gibbs.

As with the previous GP formulation we consider some posterior output. The posterior mean field is shown in Figure 48. Also shown is the residual field (true value of z - posterior mean). The RMS error here is 0.69, which is much smaller than what was obtained under the GP formulation. This is due to stronger linkage to the simulator model. This strong linkage is problematic if the simulator does not adequately model reality, in which case the GP formulation may be advantageous, even without computing considerations. Figure 49 shows the fitted values against the true field. Also shown in this figure are the pointwise 80% credible intervals for the initial field $z_0 = (z(1,0), z(2,0), \ldots, z(n_s,0))^T$. Again, these intervals are narrower as compared to those of the GP formulation and they are much closer to the true z_0.

6.4 PDE-based MRF formulation

The inverse specification is ideal if the simulator adequately matches the actual physical system being observed and if it runs quickly enough so that MCMC is possible. When this isn't the case, one is forced to replace the simulator with some sort of fast model. This necessarily will introduce additional uncertainty regarding our estimate of the true underlying field z. Here we replace the deterministic L96 simulator with MRF model whose full conditionals are constructed to match the steps in the L96 simulator model. Hence the PDE equation is replaced with a cruder, stochastic PDE (SPDE). This SPDE-based MRF model is easy to work with computationally and readily incorporates information from the observed data. The price payed here is that it is an approximation to the actual L96 system.

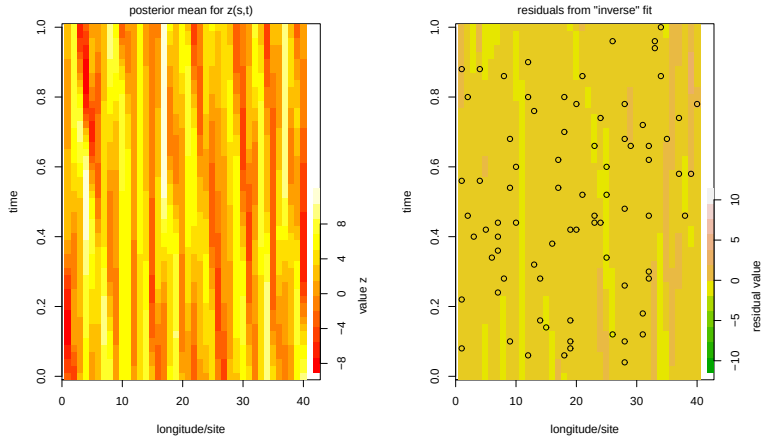


Figure 48: Left: Posterior mean for $z(s,t)$ estimated under the inverse formulation. Right: The residual field obtained by subtracting the posterior mean estimate from the true field shown in Figure 45. The circles denote locations where measurements were obtained.

For the PDE-based MRF model for the $n_s \times n_t$ field z, it is convenient to define z_t to be the n_s-vector $z_t = (z(1,t), \ldots, z(n_s,t))^T$, for $t = 0, \Delta t, \ldots, 1$. We can now model z as a Markov process

$$\pi(z|\lambda_0, \lambda_z) = \pi(z_0|\lambda_0) \times \prod_{t=1}^{n_t - 1} \pi(z_{t+\Delta t}|z_t).$$

The locally linear GMRF model previously used for z_0 is also used here.

$$\pi(z_0|\lambda_0) \propto \lambda_0^{\frac{n_s}{2}} \exp\left\{ -\tfrac{1}{2}\lambda_0 \sum_{s=1}^{n_s} (z_{s0} - z_{s+1\,0})^2 \right\}.$$

For $z_{t+\Delta t}|z_t$, the new field $z_{t+\Delta t}$ is equal to the old field z_t plus a deterministic increment $\mu_{t+\Delta t}$, plus some white noise

$$z_{t+\Delta t}|z_t = \mu_{t+\Delta t} + \delta$$

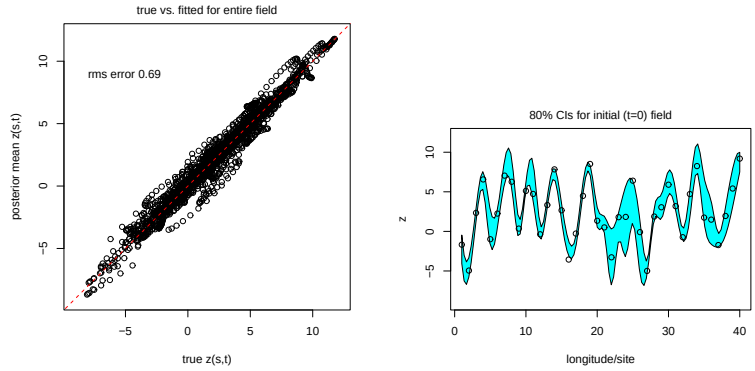


Figure 49: Left: True $z(s,t)$ plotted against the posterior mean from the inverse formulation. Right: Pointwise 80% credible intervals for the initial condition $z_0 = (z(1,0), z(2,0), \ldots, z(n_s,0))^T$. The circles denote the true values for z_0.

where $\mu_{t+\Delta t}$ is a n_s-vector obtained from the L96 step and $\delta \sim N(0, \frac{1}{\lambda_z} I_{n_s})$. Specifically, $\mu_{t+\Delta t} = (\mu_{1\,t+\Delta t}, \ldots, \mu_{40\,t+\Delta t})^T$ where

$$\mu_{s\,t+\Delta t} = z_{st} + \Delta t\{[z_{s+1\,t} - z_{s-2\,t}]z_{s-1\,t} - z_{s\,t} + F\}.$$

Thus $\pi(z_{t+\Delta t}|z_t)$ is given by

$$\pi(z_{t+\Delta t}|z_t, \lambda_z) \propto \lambda_z^{\frac{n_s}{2}} \exp\left\{-\tfrac{1}{2}\lambda_z (z_{t+\Delta t} - \mu_{t+\Delta t})^T (z_{t+\Delta t} - \mu_{t+\Delta t})\right\}.$$

The full conditional density for z_{st} can be determined by restricting attention to terms in $\pi(z|\lambda_0, \lambda_z)$ that include the

term z_{st}. Hence

$$\pi(z_{st}|z_{\partial st}, \lambda_0, \lambda_z) \propto$$
$$\exp\left\{-\frac{1}{2}\lambda_0 \sum_{j\in\partial^0 s}(z_{s0} - z_{j0})^2\right\}^{I[t=0]} \times$$
$$\exp\left\{-\frac{1}{2}\lambda_z(z_{st} - \mu_{st})^2\right\}^{I[t>0]} \times$$
$$\exp\left\{-\frac{1}{2}\lambda_y \sum_{j\in\partial^* s}(z_{j\,t+\Delta t} - \mu_{j\,t+\Delta t})^2\right\}^{I[t<1]}$$

where $\partial^0 s$ denotes the neighborhood of s for the locally linear prior for z_0, and $\partial^* s$ denotes the neighborhood of s for the MRF scheme to incorporate the PDE structure. Note that evaluating $\pi(z_{st}|z_{\partial st}, \lambda_0, \lambda_z)$ is a simple, local calculation – but nonlinear. For $0 < t < 1$, the dependence structure for z_{st} is shown in Figure 50.

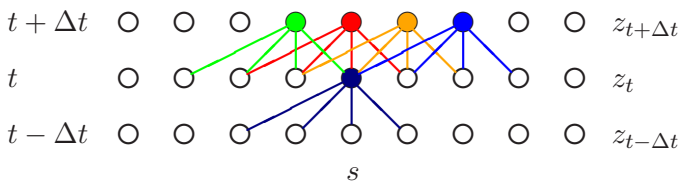


Figure 50: Dependence graph for z_{st}, a single component of the PDE-based MRF process z. z_{st}, shown by the central, dark circle, is influenced by $z_{t-\Delta t}$ at the four sites connected to z_{st}. Likewise, the four colored sites at time $t + \Delta t$ are affected by z_{st}.

Now, the PDE-based MRF formulation can be summarized as follows.

Likelihood/sampling model:

$$L(y|z) \propto \lambda_y^{\frac{n}{2}} \exp\{-\frac{1}{2}\lambda_y (y - Tz)^T (y - Tz)\}.$$

SPDE prior for z:

$$\pi(z|\lambda_0, \lambda_z) \propto \lambda_0^{\frac{p}{2}} \exp\left\{-\frac{1}{2}\lambda_0 \sum_{s=1}^{p} (z_{s0} - z_{s+1\,0})^2\right\} \times \prod_{t>0} \lambda_z^{\frac{p}{2}} \exp\left\{-\frac{1}{2}\lambda_z (z_t - \mu_t)^T (z_t - \mu_t)\right\}.$$

Priors for precision parameters:

$$\begin{aligned}
\pi(\lambda_y) &\propto \lambda_y^{a_y - 1} \exp\{-b_y \lambda_y\}, \quad a_y = 4; \quad b_y = 1 \\
\pi(\lambda_0) &\propto \lambda_0^{a_0 - 1} \exp\{-b_0 \lambda_0\}, \quad a_0 = 1; \quad b_0 = .001 \\
\pi(\lambda_z) &\propto \lambda_z^{a_z - 1} \exp\{-b_z \lambda_z\}, \quad a_z = 1; \quad b_z = .001
\end{aligned}$$

The full conditionals from this formulation are then

$$\begin{aligned}
\pi(z_{st}|\cdots) &\propto \exp\left\{-\frac{1}{2}\lambda_y (y_{st} - z_{st})^2\right\}^{I[\text{obs at } (s,t)]} \times \\
&\quad \exp\left\{-\frac{1}{2}\lambda_0 \sum_{j \in \partial^0 s} (z_{s0} - z_{j0})^2\right\}^{I[t=0]} \times \\
&\quad \exp\left\{-\frac{1}{2}\lambda_z (z_{st} - \mu_{st})^2\right\}^{I[t>0]} \times \\
&\quad \exp\left\{-\frac{1}{2}\lambda_y \sum_{j \in \partial^* s} (z_{j\,t+\Delta t} - \mu_{j\,t+\Delta t})^2\right\}^{I[t<1]} \\
\lambda_y|\cdots &\sim \Gamma(a_y + n/2, b_y + .5(y - Tz)^T (y - Tz)) \\
\lambda_0|\cdots &\sim \Gamma\left(a_0 + p/2, b_0 + .5 \sum_{i=1}^{p} (z_{0i} - z_{0i+1})^2\right) \\
\lambda_z|\cdots &\sim \Gamma\left(a_z + 50p/2, b_z + .5 \sum_{t>0} (z_t - \mu_t)^T (z_t - \mu_t)\right).
\end{aligned}$$

The MCMC implementation uses Metropolis updates for z_{st} and Gibbs updates for the precision parameters.

Again, posterior output is shown in Figures 51 and 52. Generally, the posterior reconstruction is much better than that of the GP formulation and only slightly worse than that of the inverse formulation. This performance is quite remarkable given no evaluations of the L96 simulator were carried out here.

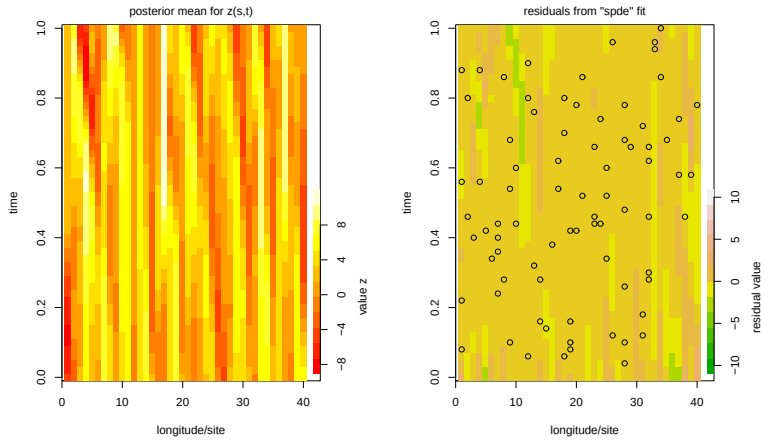


Figure 51: Left: Posterior mean for $z(s,t)$ estimated under the PDE-based MRF formulation. Right: The residual field obtained by subtracting the posterior mean estimate from the true field shown in Figure 45. The circles denote locations where measurements were obtained.

This PDE-based MRF formulation originated in geophysical applications (Wikle *et al.*, 2001; Berliner, 2003). The MRF constructed here is essentially a very coarse version of the L96 PDE system. The prior MRF by itself would not be stable if it were propagated forward since the time step is far too large. However the observed data is enough to ensure a stable reconstruction. This PDE-based MRF prior is effec-

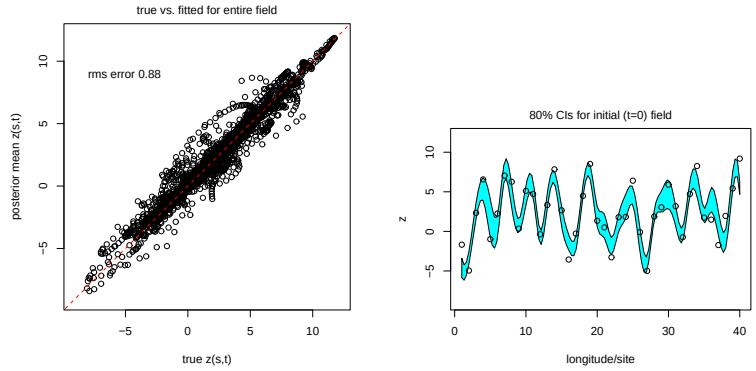


Figure 52: Left: True $z(s,t)$ plotted against the posterior mean from the PDE-based MRF formulation. Right: Pointwise 80% credible intervals for the initial condition $z_0 = (z(1,0), z(2,0), \ldots, z(n_s,0))^T$. The circles denote the true values for z_0.

tively a crude simulator for which MCMC can be carried out. For other approaches that utilize crude simulation models see Kennedy and O'Hagan (2000), Craig *et al.*(2001), or Higdon *et al.*(2003). A Kalman filtering approach for the L96 system is given in Bengtsson *et al.*(2003).

7 Discussion

This chapter introduced some basic approaches for constructing space-time models. There was a deliberate attempt here to keep the models simple so that basic modeling ideas would be a bit more clear.

One theme in the chapter has been the flexibility of Gaussian formulations. They are well developed and understood processes that offer a number of computational advantages. Even for modeling non-Gaussian processes, Gaussian systems can still serve as useful components for the eventual model.

Another facet of space-time modeling is the need to consider the details of the application at hand when developing a model formulation. Features that may require special consideration include large datasets, special dependence structure, availability of a simulation model, speed of the simulation model, and the amount of accuracy required.

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, Boca Raton.

Bengtsson, T., Snyder, C. and Nychka, D. (2003). Toward a nonlinear ensemble filter for high-dimensional systems, *Journal of Geophysical Research* **108**: 1–10.

Berliner, L. M. (2003). Physical-statistical modeling in geophysics, *Journal of Geophysical Research* **108**: D–24.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion), *Journal of the Royal Statistical Society (Series B)* **36**: 192–225.

Besag, J. (1986). On the statistical analysis of dirty pictures, *Journal of the Royal Statistical Society (Series B)* **48**: 259–302.

Besag, J. and Kooperberg, C. L. (1995). On conditional and intrinsic autoregressions, *Biometrika* **82**: 733–746.

Besag, J., Green, P. J., Higdon, D. M. and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion), *Statistical Science* **10**: 3–66.

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion), *Annals of the Institute of Statistical Mathematics* **43**: 1–59.

Chilés, J.-P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*, Wiley, New York.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association* **74**: 829–836.

Craig, P. S., Goldstein, M., Rougier, J. C. and Seheult, A. H. (2001). Bayesian forecasting using large computer models, *Journal of the American Statistical Association* **96**: 717–729.

Cressie, N. A. C. (1991). *Statistics for Spatial Data*, Wiley-Interscience.

De Oliveira, V. (2000). Bayesian prediction of clipped Gaussian random fields, *Computational Statistics & Data Analysis* **34**(3): 299–314.

Dongarra, J. J., Bunch, J. R., Moler, C. B. and Stewart, G. W. (1978). *LAPACK Users Guide*, SIAM Publications, Philadelphia.

Hastie, T., Tibshirani, R. and Friedman, J. H. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York.

Higdon, D. (1998). A process-convolution approach to modeling temperatures in the north Atlantic Ocean, *Journal of Environmental and Ecological Statistics* **5**: 173–190.

Higdon, D. (2002). Space and space-time modeling using process convolutions, *in* C. Anderson, V. Barnett, P. C. Chatwin

and A. H. El-Shaarawi (eds), *Quantitative Methods for Current Environmental Issues*, Springer Verlag, London, pp. 37–56.

Higdon, D. M., Lee, H. and Holloman, C. (2003). Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems, *in* J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West (eds), *Bayesian Statistics 7. Proceedings of the Seventh Valencia International Meeting*, Oxford University Press, pp. 181–197.

Kennedy, M. and O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available, *Biometrika* **87**: 1–13.

Lee, H. H. K., Higdon, D., Calder, K. and Holloman, C. (2005). Efficient models for correlated data via convolutions of intrinsic processes, *Statistical Modelling* **5**: 53–74.

Lorenz, E. N. (1996). Predictability, a problem partially solved, *Proceedings of the Seminar on Predictability*, Vol. 1, European Center for Medium-range Weather Forecasts, Reading, Berkshire, pp. 1–18.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21**: 1087–1091.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (2002). *Numerical Recipes in C++, The Art of Scientific Computing, Second Edition*, Cambridge University Press, New York.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Vol. 104 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London.

Rue, H. and Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian random fields, *Scandinavian Journal of Statistics* **29**: 31–49.

Ryti, R. T. (1993). Superfund soil cleanup: developing the Piazza Road remedial design, *Journal of the Air and Waste Management Association* **24**: 381–391.

Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*, Springer-Verlag, New York.

Stroud, J., Müler, P. and Sanso, B. (1999). Dynamic models for spatio-temporal data, *Technical Report 99-20*, Institute of Statistics and Decision Sciences, Duke University.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons.

Wackernagel, H. (1995). *Multivariate Geostatistics. An Introduction With Applications*, Springer-Verlag.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models (Second Edition)*, Springer-Verlag, New York.

Wikle, C. K., Milliff, R. F., Nychka, D. and Berliner, L. M. (2001). Spatio-temporal hierarchical Bayesian modeling: Tropical ocean surface winds, *Journal of the American Statistical Association* **96**: 382–397.